PCT

(51) International Patent Classification 5 :		
G11C 7/00, 11/00, 16/00 G01R 31/28	**A1**	

(11) International Publication Number: WO 90/12400

(43) International Publication Date: 18 October 1990 (18.10.90)

(72) Inventors: MEHROTRA, Sanjay ; 735 Berkshire Place, Milpitas, CA 95035 (US). HARARI, Eliyahou ; 104 Auzerais Court, Los Gatos, CA 95035 (US). LEE, Winston ; 1814 34th Avenue, San Francisco, CA 94122 (US).

(54) Title: MULTI-STATE EEPROM READ AND WRITE CIRCUITS AND TECHNIQUES

(57) Abstract

Improvements in the circuits and techniques for read, write and erase of EEprom memory (60). In the circuits for normal read, and read between write or erase for verification, the reading is made relative to set of threshold levels as provided by a corresponding set of reference cells (431, 432 etc.) which closely track and make adjustment for the variations presented by the memory cells. In one embodiment, each Flash sector of memory cells has its own reference cells for reading the cells in the sector, and a set of reference cells (529) also exists for the whole memory chip acting as a master reference.

+ See back of page
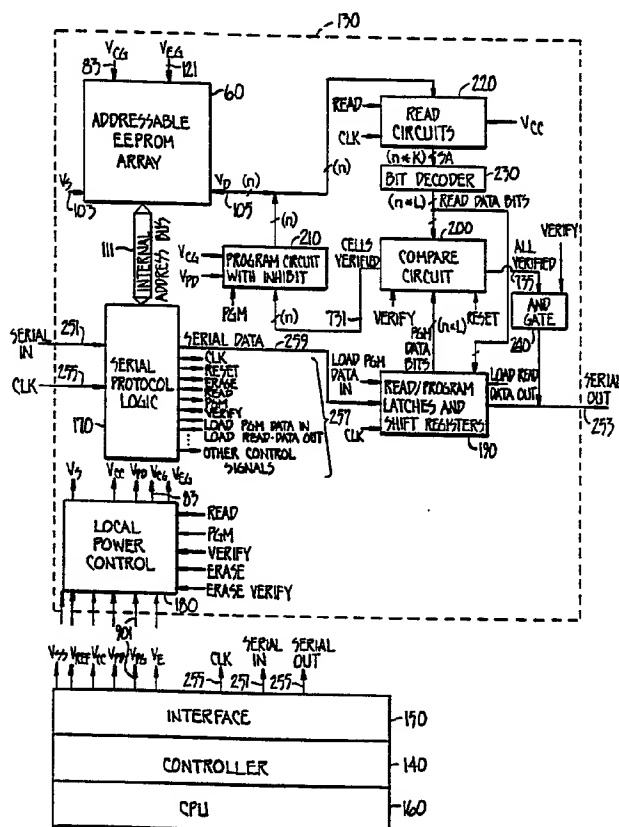
DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	HU	Hungary	NO	Norway
BJ	Benin	IT	Italy	RO	Romania
BR	Brazil	JP	Japan	SD	Sudan
CA	Canada	KP	Democratic People's Republic	SE	Sweden
CF	Central African Republic		of Korea	SN	Senegal
CG	Congo	KR	Republic of Korea	SU	Soviet Union
CH	Switzerland	LI	Liechtenstein	TD	Chad
CM	Cameroon	LK	Sri Lanka	TG	Togo
DE	Germany, Federal Republic of	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

1

MULTI-STATE EEPROM READ AND WRITE
CIRCUITS AND TECHNIQUES

5   BACKGROUND OF THE INVENTION
               This application is a continuation-in-part of
      application Serial No. 337,579 filed April 13, 1989.
               This   invention   relates   generally   to
      semiconductor electrically erasable programmable read
10   only memories (EEprom), and specifically to circuits and
      techniques for reading and programming their state.
               EEprom and electrically programmable read only
      memory (Eprom) are typically used in digital circuits
      for non-volatile storage of data or program. They can
15   be erased and have new data written or "programmed" into
      their memory cells.
               An Eprom utilizes a floating (unconnected)
      conductive gate, in a field effect transistor structure,
      positioned over but insulated from a channel region in
20   a semiconductor substrate, between source and drain
      regions.   A control gate is then provided over the
      floating gate, but also insulated therefrom.   The
      threshold voltage characteristic of the transistor is
      controlled by the amount of charge that is retained on
25   the floating gate.   That is, the minimum amount of
      voltage (threshold) that must be applied to the control
      gate before the transistor is turned "on" to permit
      conduction between its source and drain regions is
      controlled by the level of charge on the floating gate.
30   The floating gate can hold a range of charge
      and therefore an Eprom memory cell can be programmed to

2

any threshold level within a threshold window. The size
of the threshold window, delimited by the minimum and
maximum threshold levels of the device, depends on the
device's characteristics, operating conditions and
5   history.   Each distinct threshold level within the
window may, in principle, be used to designate a
definite memory state of the cell.

For Eprom memory, the transistor serving as a
memory cell is programmed to one of two states by
10   accelerating electrons from the substrate channel
region, through a thin gate dielectric and onto the
floating gate.   The memory states are erasable by
removing the charge on the floating gate by ultra-violet
radiation.

15   An electrically erasable and programmable read
only memory (EEprom) has a similar structure but
additionally provides a mechanism for removing charge
from its floating gate upon application of proper
voltages.   An array of such EEprom cells is referred to
20   as a "Flash" EEprom array when an entire array of cells,
or significant group of cells of the array, is erased
simultaneously (i.e., in a flash).   Once erased, a cell
can then be reprogrammed.

A specific, single cell in a two-dimensional
25   array of Eprom, EEprom cells is addressed for reading by
application of a source-drain voltage to source and
drain lines in a column containing the cell being
addressed, and application of a control gate voltage to
a word line connected to the control gates in a row
30   containing the cell being addressed.

An addressed memory cell transistor's state is
read by placing an operating voltage across its source
and drain and on its control gate, and then detecting
the level of current flowing between the source and
35   drain.   The level of current is proportional to the

3

threshold level of the transistor, which in turn is determined by the amount of charge on its floating gate.

In the usual two-state EEprom cell, one breakpoint threshold level is established so as to
5   partition the threshold window into two regions. The source/drain current is compared with the breakpoint threshold level that was used when the cell was programmed. If the current read is higher than that of the threshold, the cell is determined to be in a "zero"
10   state, while if the current is less than that of the threshold, the cell is determined to be in the other state. Thus, such a two-state cell stores one bit of digital information. A current source which may be externally programmable is often provided as part of a
15   memory system to generate the breakpoint threshold current.

Thus, for a multi-state EEprom memory cell, each cell stores two or more bits of data. The information that a given EEprom array can store is thus
20   increased by the multiple of number of states that each cell can store.

Accordingly, it is a primary object of the present invention to provide a system of EEprom memory cells wherein the cells are utilized to store more than
25   one bit of data.

It is a further object of the present invention to provide improved read circuits as part of an Eprom or EEprom integrated circuit memory chip.

It is also an object of the invention to
30   provide read circuits which are simpler, easier to manufacture and have improved accuracy and reliability over an extended period of use.

It is also an object of the present invention to provide improved program circuits as part of an Eprom
35   or EEprom integrated circuit memory chip.

4

It is also an object of the invention to provide program circuits which are simpler, easier to manufacture and have improved accuracy and reliability over an extended period of use.

It is another object of the present invention to provide memory read and program techniques that automatically compensate for effects of temperature, voltage and process variations, and charge retention.

It is yet another object of the present invention to provide Flash EEprom semiconductor chips that can replace magnetic disk storage devices in computer systems.

Further, it is an object of the present invention to provide a Flash EEprom structure capable of an increased lifetime as measured by the number of program/read cycles that the memory can endure.

SUMMARY OF THE INVENTION

These and additional objects are accomplished by improvements in EEprom array read and write circuits and techniques in order to provide multiple threshold levels that allow accurate reading and writing of more than two distinct states within each memory cell over an extended lifetime of the memory cells, so that more than one bit may be reliably stored in each cell.

According to one aspect of the present invention, the multiple threshold breakpoint levels are provided by a set of memory cells which serves as master reference cells. The master reference cells are independently and externally programmable, either by the memory manufacturer or the user. This feature provides maximum flexibility, allowing the breakpoint thresholds to be individually set within the threshold window of the device at any time. Also, by virtue of being an identical device as that of the memory cells, the reference cells closely track the same variations due to

5

manufacturing processes, operating conditions and device aging. The independent programmability of each breakpoint threshold level allows optimization and fine-tuning of the threshold window's partitioning, critical in multi-state implementation. Furthermore, it allows post-manufacture configuration for either 2-state or multi-state memory from the same device, depending on user need or device characteristics at the time.

According to another aspect of the present invention, a set of memory cells within each sector (where a sector is a group of memory cells which are all erased at the same time in a Flash EEprom) are set aside as local reference cells. Each set of reference cells tracks the Flash cells in the same sector closely as they are both cycled through the same number of program/erase cycles. Thus, the aging that occurs in the memory cells of a sector after a large number of erase/reprogram cycles is also reflected in the local reference cells. Each time the sector of flash cells is erased and reprogrammed, the set of individual breakpoint threshold levels are re-programmed to the associated local reference cells. The threshold levels read from the local reference cells then automatically adjust to changing conditions of the memory cells of the same sector. The threshold window's partitioning is thus optimally maintained. This technique is also useful for a memory that employs only a single reference cell that is used to read two state (1 bit) memory cells.

According to another aspect of the present invention, the threshold levels rewritten at each cycle to the local reference cells are obtained from a set of master cells which are not cycled along with the memory cells but rather which retain a charge that has been externally programmed (or reprogrammed). Only a single

6

set of master memory cells is needed for an entire
memory integrated circuit.

In one embodiment, the read operation directly
uses the threshold levels in the local reference cells

5   previously copied from the master reference cells. In
another embodiment, the read operation indirectly uses
the threshold levels in the local reference cells even
though the reading is done relative to the master
reference cells. It does this by first reading the

10 local reference cells relative to the master reference
cells. The differences detected are used to offset
subsequent regular readings of memory cells relative to
the master reference cells so that the biased readings
are effectively relative to the local reference cells.

15   According to another aspect of the present
invention, a read operation on a memory cell determines
which memory state it is in by comparing the current
flowing therethrough with that of a set of reference
currents corresponding to the multiple threshold

20 breakpoint levels.

In one embodiment, the current flowing through
a cell being read is compared one-by-one with each of
the threshold current levels of the reference cells.

In another embodiment, the current flowing

25 through a cell to be read is compared simultaneously
with that of the set of reference cells. A special
current mirror configuration reproduces the current to
be read without degrading its signal, into multiple
branches, one for each threshold current comparison.

30   According to another aspect of the present
invention, the program and verify operations are
performed on a chunk (i.e. several bytes) of addressed
cells at a time. Furthermore, the verify operation is
performed by circuits on the EEprom chip. This avoids

35 delays in shipping data off chip serially for
verification in between each programming step.

7

According to another aspect of the present invention, where a programmed state is obtained by repetitive steps of programming and verifying from the "erased" state, a circuit verifies the programmed state after each programming step with the intended state and selectively inhibits further programming of any cells in the chunk that have been verified to have been programmed correctly. This enables efficient parallel programming of a chunk of data in a multi-state implementation.

According to another aspect of the present invention, where a chunk of EEprom cells are addressed to be erased in parallel, an erased state is obtained by repetitive steps of erasing and verifying from the existing state to the "erased" state, a circuit verifies the erased state after each erasing step with the "erased" state and selectively inhibits further erasing of any cells in the chunk that have been verified to have been erased correctly. This prevents over-erasing which is stressful to the device and enables efficient parallel erasing of a group of cells.

According to another aspect of the present invention, after a group of cells have been erased to the "erased" state, the cells are re-programmed to the state adjacent the "erased" state. This ensures that each erased cell starts from a well defined state, and also allows each cell to undergo similar program/erase stress.

According to another aspect of the present invention, the voltages supplied to the control gates of the EEprom cells are variable over a wide range and independent of the voltage supplied to the read circuits. This allows accurate program/erase margining as well as use in testing and diagnostics.

The subject matter herein is a further development of the EEprom array read techniques

8

described in copending patent application Serial No.
204,175, filed June 8, 1988, by Dr. Eliyahou Harari,
particularly the disclosure relating to Figure 11e
thereof.   Application Serial No. 204,175 is hereby

5 expressly incorporated herein by reference, the
disclosure with respect to the embodiments of Figures
11, 12, 13 and 15 being most pertinent.

         Additional objects, features and advantages of
the present invention will be understood from the

10 following description of its preferred embodiments,
which description should be taken in conjunction with
the accompanying drawings.

               BRIEF DESCRIPTION OF THE DRAWINGS
               Figure 1 is a cross-sectional view of an

15 EEprom device integrated circuit structure that can be
used to implement the various aspects of the present
invention;

               Figure 2 is a view of the structure of Figure
1 taken across section 2-2 thereof;

20   Figure 3 is an equivalent circuit of a single
EEprom cell of the type illustrated in Figures 1 and 2;

               Figure 4 shows an addressable array of EEprom
cells;

               Figure 5 is a block diagram of an EEprom

25 system in which the various aspects of the present
invention are implemented;

               Figure 6 illustrates the partitioning of the
threshold window of an EEprom cell which stores one bit
of data;

30   Figure 7A illustrates the partitioning of the
threshold window of an EEprom cell which stores two bits
of data;

               Figure 7B illustrates the partitioning of the
source-drain conduction current threshold window of the

35 EEprom cell of figure 7A;

9

Figures 8A and 8B are curves that illustrate the changes and characteristics of a typical EEprom after a period of use;

Figure 9A illustrates read and program
5   circuits for a master reference cell and an addressed memory cell according to the present invention;

Figure 9B illustrates multi-state read circuits with reference cells according to the present invention;

10   Figures 9C(1)-9C(8) illustrate the timing for multi-state read for the circuits of Figure 9B;

Figure 9D illustrates one embodiment of a multi-state read circuit in which the memory state of an address cell is sensed relative to a set of reference
15   current levels simultaneously;

Figure 9E illustrates one embodiment of an IREF circuit shown in Figure 9D as an EEprom cell programmed with a reference current;

Figure 9F illustrates a preferred
20   implementation of the embodiment in Figure 9D in which each IREF circuit is provided by a current source reproducing a reference current programmed in the EEprom cell;

Figure 9G illustrates another embodiment of an
25   IREF circuit shown in Figure 9D in which a reference current is provided in each branch by the conduction of a transistor of predetermined size;

Figure 9H illustrates another embodiment of a multi-state read circuit in which the memory state of an
30   address cell is sensed relative to a set of reference current levels simultaneously;

Figure 9I illustrates yet another embodiment of a multi-state read circuit in which the memory state of an address cell is sensed relative to a set of
35   reference current levels simultaneously;

10

Figure 10 illustrates a specific memory organization according to the present invention;

Figure 11 shows an algorithm for programming a set of local reference cells according to the present invention;

Figure 12A shows one embodiment of a read circuit using local reference cells directly;

Figure 12B shows a read algorithm for the embodiment of Figure 12A;

Figure 13A shows an alternative embodiment of a read circuit using local reference cells indirectly;

Figure 13B is a programmable circuit for the biased reading of the master reference cells according the alternative embodiment;

Figure 13C is a detail circuit diagram for the programmable biasing circuit of Figure 13B;

Figure 13D shows a read algorithm for the embodiment of Figure 13A;

Figure 14 illustrates the read/program data paths for a chunk of cells in parallel;

Figure 15 shows an on chip program/verify algorithm according to the present invention;

Figure 16 is a circuit diagram for the compare circuit according to the present invention;

Figure 17 is a circuit diagram for the program circuit with inhibit according to the present invention;

Table 1 and 2 list typical examples of operating voltages for the EEprom cell of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

There are many specific Eprom, EEprom semiconductor integrated circuit structures that can be utilized in making a memory array with which the various aspects of the present invention are advantageously implemented.

11

<u>"Split-Channel" EEprom Cell</u>

A preferred EEprom structure is generally
illustrated in the integrated circuit cross-sectional
views of Figures 1 and 2. Describing this preferred
5   structure briefly, two memory cells 11 and 13 are formed
on a lightly p-doped substrate 15. A heavily n-doped
implanted region 17 between the cells 11 and 13 serves
as a drain for the cell 11 and a source for the cell 13.
Similarly, another implanted n-doped region 19 is the
10   source of the cell 11 and the drain of an adjacent cell,
and similarly for another n-doped region 21.

Each of the memory cells 11 and 13 contains
respective conductive floating gates 23 and 25,
generally made of polysilicon material. Each of these
15   floating gates is surrounded by dielectric material so
as to be insulated from each other and any other
conductive elements of the structure. A control gate 27
extends across both of the cells 11 and 13 in a manner
to be insulated from the floating gates and the
20   substrate itself. As shown in Figure 2, conductive
strips 29 and 31 are additionally provided to be
insulated from each other and other conductive elements
of the structure, serving as erase gates. A pair of
such erase gates surrounds the floating gate of each
25   memory cell and are separated from it by an erase
dielectric layer. The cells are isolated by thick field
oxide regions, such as regions 33, 35, and 37, shown in
the cross-section of Figure 1, and regions 39 and 41
shown in the view of Figure 2.

30   The memory cell is programmed by transferring
electrons from the substrate 15 to a floating gate, such
as the floating gate 25 of the memory cell 13. The
charge on the floating gate 25 is increased by electrons
travelling across the dielectric from a heavily p-doped
35   region 43 and onto the floating gate. Charge is removed
from the floating gate through the dielectric between it

12

and the erase gates 29 and 31. This preferred EEprom
structure, and a process for manufacturing it, are
described in detail in copending patent application
Serial No. 323,779 of Jack H. Yuan and Eliyahou Harari,
5   filed March 15, 1989, which is expressly incorporated
herein by reference.

The EEprom structure illustrated in Figures 1
and 2 is a "split-channel" type. Each cell may be
viewed as a composite transistor consisting of two
10   transistor T1 and T2 in series as shown in Figure 3. The
T1 transistor 11a is formed along the length L1 of the
channel of the cell 11 of Figure 1. It has a variable
threshold voltage V_{T1}. In series with the T1 transistor
11a is the T2 transistor 11b that is formed in a portion
15   of the channel L2. It has a fixed threshold voltage V_{T2}
of about 1V. Elements of the equivalent circuit of
Figure 3 are labeled with the same reference numbers as
used for corresponding parts in Figures 1 and 2, with a
prime (') added.

20   As can best be seen from the equivalent
circuit of Figure 3, the level of charge on the T1's
floating gate 23' of an EEprom cell affects the
threshold voltage V_{T1} of the T1 transistor 11a when
operated with the control gate 27'. Thus, a number of
25   memory states may be defined in a cell, corresponding to
well defined threshold voltages programmed into the cell
by appropriate amount of charges placed on the floating
gate. The programming is performed by applying, over a
certain period of time, appropriate voltages to the
30   cell's control gate 27' as well as drain 17' and source
19'.

Addressable Flash EEprom Array

The various aspects of the present invention
are typically applied to an array of Flash EEprom cells

13

in an integrated circuit chip. Figure 4 illustrates
schematically an array of individually addressable
EEprom cells 60. Each cell is equivalent to the one
shown in Figure 3, having a control gate, source and
5   drain, and an erase gate. The plurality of individual
memory cells are organized in rows and columns. Each
cell is addressed by selectively energizing its row and
column simultaneously. A column 62, for example,
includes a first memory cell 63, an adjacent second
10   memory cell 65, and so forth. A second column 72
includes memory cells 73, 75, and so forth. Cells 63
and 73 are located in a row 76, cells 65 and 71 in
another, adjacent row, and so forth.

Along each row, a word line is connected to
15   all the control gates of the cells in the row. For
example, the row 76 has the word line 77 and the next
row has the word line 79. A row decoder 81 selectively
connects the control gate voltage V_{CG} on an input line
83 to all the control gates along a selected word line
20   for a row.

Along each column, all the cells have their
sources connected by a source line such as 91 and all
their drains by a drain line such as 93. Since the
cells along a row are connected in series by their
25   sources and drains, the drain of one cell is also the
source of the adjacent cell. Thus, the line 93 is the
drain line for the column 62 as well as the source line
for the column 72. A column decoder 101 selectively
connects the source voltage V_S on an input line 103 to
30   all the sources and connects the drain voltage V_D on an
input line 105 to all the drains along a selected
column.

Each cell is addressed by the row and column
in which it is located. For example, if the cell 75 is
35   addressed for programming or reading, appropriate
programming or reading voltages must be supplied to the

14

cell's control gate, source and drain. An address on
the internal address bus 111 is used to decode row
decoder 81 for connecting V_{CG} to the word line 79
connected to the control gate of the cell 75. The same
5   address is used to decode column decoder 101 for
connecting V_S to the source line 93 and V_D to the drain
line 95, which are respectively connected to the source
and drain of the cell 75.

        One aspect of the present invention, which
10   will be disclosed in more detail in a later section, is
the implementation of programming and reading of a
plurality of memory cells in parallel. In order to
select a plurality of columns simultaneously, the column
decoder, in turn, controls the switching of a source
15   multiplexer 107 and a drain multiplexer 109. In this
way, the selected plurality of columns may have their
source lines and drain lines made accessible for
connection to V_S and V_D respectively.

        Access to the erase gate of each cell is
20   similar to that of the control gate. In one
implementation, an erase line such as 113 or 115 or 117
is connected to the erase gate of each cells in a row.
An erase decoder 119 decodes an address on the internal
address bus 111 and selectively connects the erase
25   voltage V_{EG} on input line 121 to an erase line. This
allows each row of cells to be addressed independently,
such as the row 76 being simultaneously (Flash) erased
by proper voltages applied to their erase gates through
erase line 113. In this case, the Flash cell consists
30   of one row of memory cells. However, other Flash cell's
implementations are possible and most applications will
provide for simultaneous erasing of many rows of cells
at one time.

Flash EEprom System

15

The addressable EEprom array 60 in figure 4 forms part of the larger multi-state Flash EEprom system of the present invention as illustrated in figure 5. In the larger system, an EEprom integrated circuit chip 130
5   is controlled by a controller 140 via an interface 150. The controller 140 is itself in communication with a central microprocessor unit 160.

The EEprom chip 130 comprises the addressable EEprom array 60, a serial protocol logic 170, local
10 power control circuits 180, and various programming and reading circuits 190, 200, 210, 220, 230 and 240.

The controller 140 controls the functioning of the EEprom chip 130 by supplying the appropriate voltages, controls and timing. Tables 1 and 2 shows
15 typical examples of voltage conditions for the various operational modes of the EEprom cell. The addressable EEprom array 60 may be directly powered by the controller 140 or, as shown in figure 5, be further regulated on chip by the local power control 180.
20 Control and data linkages between the controller 140 and the chip 130 are made through the serial in line 251 and the serial out line 253. Clock timing is provided by the controller via line 255.

In a typical operation of the EEprom chip 130,
25 the controller 140 will send a serial stream of signals to the chip 130 via serial in line 251. The signals, containing control, data, address and timing information, will be sorted out by the serial protocol logic 170. In appropriate time sequence, the logic 170
30 outputs various control signals 257 to control the various circuits on the chip 130. It also sends an address via the internal address bus 111 to connect the addressed cell to voltages put out from the controller. In the meantime, if the operation is programming, the
35 data is staged for programming the addressed cell by

16

being sent via a serial data line 259 to a set of
read/program latches and shift registers 190.

Read Circuits and Techniques Using Reference Cells

5   To accurately and reliably determine the
memory state of a cell is essential for EEprom
operations. This is because all the basic functions
such as read, erase verify and program verify depend on
it. Improved and novel read circuits 220 for the EEprom
chip 130 and techniques of the present invention make
10   multi-state EEprom feasible.

       As discussed in connection with figure 3, the
programmed charge placed on the floating gate 23'
determines the programmed threshold voltage V_{T1} of the
cell. Generally, V_{T1} increases or decreases with the
15   amount of negative charge on the floating gate 23'. The
charge can even be reduced to a positive value
(depletion mode) where V_{T1} decreases below V_{T2} and even
becomes negative. The maximum and minimum values of V_{T1}
are governed by the dielectric strength of the device
20   material. The span of V_{T1} defines a threshold voltage
window in which memory states may be implemented.

       Copending patent application Serial No.
204,175, discloses an EEprom cell with memory states
defined within a maximized window of threshold voltage
25   V_{T1}. The full threshold voltage window includes the
negative region of the threshold voltage, in addition to
the usual positive region. The increased window
provides more memory space to implement multi-state in
an EEprom cell.

30   Figures 6 and 7 respectively illustrates the
manner in which the threshold voltage window is
partitioned for a 2-state memory and a 4-state memory
cell. (Of course it is also possible to partition the
window for a 3-state memory or even for a continuum of
35   states in an analog, rather than digital memory).

17

Referring first to figure 6, the solid curve 343 shows V_{T1} as a function of programming time. The threshold voltage window is delimited by the minimum and maximum values of V_{T1}, represented approximately by the

5   Erase state level 345 and the Fully Program state level 347 respectively. The 2-state memory is implemented by partitioning the window into two halves 346, 348 using a breakpoint threshold level 349. Thus, the cell may be considered to be in memory state 0 (or state 1) if the

10   cell is programmed with a V_{T1} within region 346 (or region 348) respectively.

A typical erase/program cycle begins with erase which reduces the threshold voltage of the cell to its Erase state level 345. Subsequent repetitive

15   programming is used to increase the threshold voltage V_{T1} to the desired level. Rather than continuously applying programming voltages to the addressed cell for some fixed period of time corresponding to the state to which the cell is to be programmed, it is preferable to apply

20   programming voltages in repetitive short pulses with a read operation occurring after each pulse to determine when it has been programmed to the desired threshold voltage level, at which time the programming terminates. The programming voltages and duration of the pulses are

25   such that the pulses advance V_{T1} across the various regions rapidly but each pulse is sufficiently fine to not overshoot any of the regions. This minimizes voltage and field related stresses on the cell, and therefore improves its reliability.

30   Figure 7A illustrates the 4-state case where the threshold voltage window is partitioned into four regions 351, 353, 355, 357 by breakpoint levels 352, 354, 356 respectively. The cell is considered to be in state "3" or "2" or "1" or "0" if its V_{T1} is programmed

35   to be within corresponding regions 351 or 353 or 355 or 357 respectively. A 4-state cell is able to store two

18

bits of data. Thus, the four states may be encoded as
(1,1), (1,0), (0,1) and (0,0) respectively.

In general, if each EEprom cell is to store K
states, the threshold window must be partitioned into K

5   regions with at least K-1 threshold levels. Thus, only
one breakpoint level is required for a 2-state memory
cell, and three breakpoint levels are required for a 4-
state cell.

In principle, a threshold voltage window may

10   be partitioned to a large number of memory states. For
example, for an EEprom device with a maximum threshold
window of 16V, it may be partitioned into thirty-two
states each within an approximately half volt interval.
In practice, prior art EEprom devices have only stored

15   two states or one bit per cell with diminished
reliability and life. Apart from operating with a
smaller threshold window, prior devices fail to solve
two other problems inherent in EEprom devices. Both
problems relate to the uncertainty in the amount of

20   charge in the floating gate and hence the uncertainty in
the threshold voltage V_{T1} programmed into the cell.

The first problem has to do with the
endurance-related stress the device suffers each time it
goes through an erase/program cycle. The endurance of

25   a Flash EEprom device is its ability to withstand a
given number of program/erase cycles. The physical
phenomenon limiting the endurance of prior art Flash
EEprom devices is trapping of electrons in the active
dielectric films of the device. During programming,

30   electrons are injected from the substrate to the
floating gate through a dielectric interface.
Similarly, during erasing, electrons are extracted from
the floating gate to the erase gate through a dielectric
interface. In both cases, some of the electrons are

35   trapped by the dielectric interface. The trapped
electrons oppose the applied electric field in

19

subsequent program/erase cycles thereby causing the programmed V_{T1} to shift to a lower value and the erased V_{T1} to shift to a higher value. This can be seen in a gradual closure in the voltage "window" between the "0" and "1" states of prior art devices as shown in figure 8A. Beyond approximately 1×10^4 program/erase cycles the window closure can become sufficiently severe to cause the reading circuitry to malfunction. If cycling is continued, the device eventually experiences catastrophic failure due to a ruptured dielectric. This typically occurs at between 1×10^6 and 1×10^7 cycles, and is known as the intrinsic breakdown of the device. In prior art EEprom devices the window closure is what limits the practical endurance to approximately 1×10^4 program/erase cycles. This problem is even more critical if multi-state memory is implemented, since more accurate placement of V_{T1} is demanded.

A second problem has to do with the charge retention on the floating gate. The charge on the floating gate tends to diminish somewhat through leakage over a period of time. This causes the threshold voltage V_{T1} to shift also to a lower value over time. Figure 8B illustrates the reduction of V_{T1} as a function of time. Over the life time of the device V_{T1} may shift by as much as 1V. In a multi-state device, this could shift the memory by one or two states.

The present invention overcomes these problems and presents circuits and techniques to reliably program and read the various states even in a multi-state implementation.

The memory state of a cell may be determined by measuring the threshold voltage V_{T1} programmed therein. Alternatively, as set forth in co-pending patent application, Serial No. 204,175, the memory state may conveniently be determined by measuring the differing conduction in the source-drain current I_{DS} for

20

the different states. In the 4-state example, figure 7A
shows the partition in the threshold voltage window.
Figure 7B, on the other hand, illustrates typical values
of I_{DS} (solid curves) for the four states as a function

5   of the control gate voltage V_{CG}. With V_{CG} at 5V, the I_{DS}
values for each of the four conduction states can be
distinguished by sensing with four corresponding current
sensing amplifiers in parallel. Associated with each
amplifier is a corresponding reference conduction states

10   I_{REF} level (shown as broken curves in figure 8). Just as
the breakpoint threshold levels (see figures 6 and 7A)
are used to demarcate the different regions in the
threshold voltage window, the I_{REF} levels are used to do
the same in the corresponding source-drain current

15   window. By comparing with the I_{REF}'s, the conduction
state of the memory cell can be determined. Co-pending
patent application, Serial No. 204,175 proposes using
the   same sensing amplifiers and I_{REF}'s for both
programming and reading. This provides good tracking

20   between the reference levels (broken curves in figure
89) and the programmed levels (solid curves in figure
7B).

In the improved scheme of the present
invention, the I_{REF}'s are themselves provided by the

25   source-drain currents of a set of EEprom cells existing
on the same chip and set aside solely for this purpose.
Thus, they act as master reference cells with their
I_{REF}'s used as reference levels for the reading and
programming of all other EEprom cells on the same chip.

30   By using the same device as the EEprom cells to act as
reference cells, excellent tracking with respect to
temperature, voltage and process variations is achieved.
Furthermore, the charge retention problem, important in
multi-state implementation, is alleviated.

35   Referring to figure 9A, one such master
reference cell 400 is shown with its program and read

21

paths. The reference cells erase and program module 410
serves to program or re-program each such reference cell
400. The module 410 includes program and erase circuits
411 with a programming path 413 connected to the drain

5   of the master reference cell 400. The circuits 411 are
initiated by addresses decoded from the internal bus 111
by a program decoder 415 and an erase decoder 417
respectively. Accordingly, programming voltages or
erasing voltages are selectively supplied each reference

10   cell such as cell 400. In this way, the reference
level in each reference cell may be independently set or
reprogrammed. Typically, the threshold level of each
reference cell will be factory-programmed to the optimum
level appropriate for each batch of chips produced.

15   This could be done by comparison with an external
standard reference level. By software control, a user
also has the option to reset the reference threshold
levels.

       Once the reference threshold voltage V_{T1} or

20   reference drain-source current I_{REF} is programmed into
each reference cell 400, it then serves as a reference
for the reading of an addressed memory cell such as cell
420. The reference cell 400 is connected to a first leg
403 of a current sensing amplifier 410 via a clocked

25   switch 413. A second leg 415 of the amplifier is
essentially connected to the addressed memory cell 420
whose programmed conduction state is to be determined.
When cell 420 is to be read, a control signal READ will
enable a switch 421 so that the cell's drain is

30   connected to the second leg 415. The sense amplifier
410 supplies voltage via V_{CC} to the drains of both the
master reference cell 400 and the addressed cell 420.
In the preferred embodiment, the amplifier has a current
mirror configuration such that any differential in

35   currents through the two legs 403 and 415 results in the
voltage in the second leg 415 being pulled up towards V_{CC}

22

or down towards V_S. Thus, the node at the second leg
415 is respectively HIGH (or LOW) when the source-drain
current I_{DS} for the addressed cell 420 is less (or more)
than I_{REF} through the master reference cell 400. At the

5   appropriate time controlled by a clocked switch 423, the
sensed result at the second leg 415 may be held by a
latch 425 and made available at an output line 427.
When I_{DS} is less than I_{REF}, a HIGH appears at the output
line 427 and the addressed cell 420 is regarded as in

10   the same conduction state as the master reference cell
400.

In the preferred embodiment, a voltage clamp
and fast pull-up circuit 430 is also inserted between
the second leg 415 and the drain 431 of the addressed

15   cell 420. The circuit 430 serves to keep the drain
voltage V_D at a maximum of 1.5V - 2.0V when it is
charging up in the case of lower I_{DS}. It also prevents
V_D from pulling too low in the case of higher I_{DS}.

In general, if each memory cell is to store K

20   states, then at least K-1, or preferably K reference
levels need be provided. In one embodiment, the
addressed cell is compared to the K reference cells
using k sense amplifiers in parallel. This is
preferable for the 2-state case because of speed, but

25   may spread the available current too thin for proper
sensing in the multi-state case. Thus, for multi-state
case, it is preferable to compare the addressed cell
with the K reference cells one at a time in sequence.

Figure 9B illustrates more explicitly the

30   multi-state reading configuration. The K reference
cells such as 431, 433, 435 are connected to the sense
amplifier 440 via the amplifier's first leg 441. The
connection is time-multiplexed by clocked switches such
as 451, 453, 455 respectively. The second leg 457 of

35   the sense amplifier is connected to the addressed cell
as in figure 9A. The sensed signal at the second leg

23

457 is time-selectively latched by clocked switches such as 461, 463, 465 onto such latches 471, 473, 475.

Figures 9C(1)-9C(8) illustrates the timing for multi-state read. When the signal READ goes HIGH, a switch 421 is enabled and the addressed memory cell is connected to the second leg 457 of the sense amplifier 440 (figure 9C(1)). The clocks' timing is given in figures 9C(2)-9C(4). Thus, at each clock signal, the sense amplifier sequentially compares the addressed cell with each of the reference cells and latches each results. The latched outputs of the sense amplifier are given in figures 9C(5)-9C(7). After all the K output states of the sense amplifier 440 are latched, they are encoded by a K-L decoder 480 ($2^L \geq K$) (figure 9C(8)) into L binary bits.

Thus, the multiple threshold levels are provided by a set of memory cells which serves as master reference cells. The master reference cells are independently and externally erasable and programmable, either by the device manufacturer or the user. This feature provides maximum flexibility, allowing the breakpoint thresholds to be individually set within the threshold window of the device at any time. By virtue of being the same device as that of the memory cells, the reference cells closely track the same variations due to manufacturing processes, operating conditions and charge retention problems. The independent programmability of each threshold level at will allows optimization and fine-tuning of the partitioning of the threshold window to make multi-state memory viable. Furthermore, it allows post-manufacture configuration for either 2-state or multi-state memory from the same device, depending on user need or device characteristics at the time.

Another aspect of the present invention provides improved multi-state sensing of an addressed

24

memory cell. As discussed in connection with an earlier
embodiment for sensing a mult-state memory, it is
preferable to compare the cell's conduction current with
all the reference conduction current levels (threshold
5   levels) simultaneously or in parallel. For example, a
4-state memory cell has at least three reference current
levels to demarcate the four states. Parallel sensing
the state of the cell means simultaneous comparison of
the cell's conduction current I_{CELL} versus each of the
10 three reference current levels. This is faster than
comparing with each of the three reference conduction
levels sequentially. However, in the simpler embodiment
described earlier, the conduction current of the
addressed cell would be diluted by being divided up into
15 three branches, one for each reference level comparison.
Thus, a simple implementation of simultaneous or
parallel multi-state sensing may be prohibited by the
signal-to-noise ratio requirement of the sensing system,
especially when there are many states involved.
20   Figure 9D - Figure 9I illustrate several
embodiments of simultaneous multi-state sensing without
the disadvantage of degrading the conduction current of
the sensed cell. In each embodiment, a one-to-many
current mirror is employed to reproduce a current into
25 many copies so that each copy may be used to compare
with a reference current level at the same time.
            Figure 9D illustrates a first embodiment of
simultaneous multi-state sensing. A one-to-many current
mirror comprises a first transistor 910 on a first leg
30 920 and a second transistor 911, 912, ..., 915
respectively on each branch 921, 922, ..., 925 of a
second leg. Whenever, a first current flows in the
first leg 920, the second transistor on each branch of
the second leg behaves as a current source and supplies
35 a reproduced current in its branch. The ratio of
reproduced current to the first current scales according

25

to the relative sizes of the second transistor 911, 912,
..., 915 to the first transistor 910.

In the present embodiment, all the transistors
have the same size as denoted by the symbol "X" shown in
5   Figure 9D. This results in a one-to-many current mirror
in which the first current in the first leg 920 is
identically reproduced in all branches 921, 922, ...,
925 of the second leg. Thus, when the conduction
current I_{CELL} of an addressed memory cell 420 flows
10   through a read enabling switch 421 in the first leg 920,
the same current I_{CELL} is reproduced in the branches 921,
922, ..., 925 of the second leg. This is achieved
without dilution of I_{CELL}.

Once I_{CELL} is reproduced in each branch, it is
15   compared to an associated reference current level. This
is accomplished by also driving each branch with a
second current source 931, 932, ..., 935 in-line with
the first current source 911, 912, ..., 915
respectively. Each second current source or I_{REF} circuit
20   931, 932, ..., 935 supplies respectively the
predetermined reference current level such as I_{REF1} in
line 941 of the first branch, I_{REF2} in line 942 of the
second branch, ..., and I_{REFk} in line 953 of the kth
branch. The memory state is then determined by sensing
25   the location of the I_{CELL} level relative to the I_{REF}'s.
The sensed outputs for each state denoted by SA1, SA2,
..., SAk in Figure 9D are respectively derived from a
node 951 of the first branch, a node 952 of second
branch, ..., and a node 953 of the kth branch. The node
30   in each branch is situated between the first and second
current source. In general the two current sources are
of opposite polarity. If the second current source 931,
932, ..., 935 is an n-channel transistor connected to V_s
on one end, then the first current source is a p-channel
35   transistor 911, 912, ..., 915 connected to V_{cc} on the
other end. Depending on the relative levels of I_{CELL} and

26

I_{REF} in the two current sources, the node is either pulled up towards V_{CC} (typically, 5V) or down towards V_S (typically, 0V). For example, in the first branch, a current I_{CELL} is reproduced in line 921 and a current I_{REF1}

5 is supplied in line 941. The node 951 is respectively HIGH (or LOW) when I_{CELL} is greater than (or less than) I_{REF1}. Thus, a memory state having an I_{CELL} that lies between I_{REF1} and I_{REF2} would only have the node 951 HIGH, thereby resulting in a multi-state output (SA1, SA2,

10 ..., SAk) = (0, 1, ..., 1).

In general, each I_{REF} circuit 931, 932, ..., 935 can be a current source circuit pre-adjusted to supply the various reference current levels I_{REF1}, I_{REF2}, ..., I_{REF3}.

15 Figure 9E illustrates one embodiment in EEprom applications in which each I_{REF} circuit 931, 932, ..., 935 is provided respectively by a reference cell 431, 432, ..., 435 which is itself an EEprom cell similar to that described in connection with Figures 9A and 9B.

20 Thus the reference cell may be applicable as a master reference cell or a local reference cell in which a reference conduction current level may be programmed.

Figure 9F illustrates a preferred implementation where each I_{REF} circuit is not provided

25 directly by a reference cell, but rather by a reproduction of it. This enables a chunk (e.g., 64) of memory cells to share the same reference cell for simultaneous sensing. A transistor 961, 962, ..., 965 respectively in each of the I_{REF} circuit 931, 932, ...,

30 935 serves as a current source for supplying the reproduced reference current level from each of the reference cells 431, 432, ..., 435. Each transistor is controlled by a reference voltage REF1, REF2, ..., REFk at its gate to produce the required reference current

35 levels I_{REF1}, I_{REF2}, ..., I_{REF3}. Each reference voltage is furnished by a REF circuit 971, ..., ..., 975. An

27

alternative view is that each transistor 961, 962, ...,
965 and the associated REF circuit 971, ..., ..., 975
form a double current mirror circuit by which the
reference current of each reference cell 431, 432, ...,
5   435 is reproduced as the conduction current of the
transistor 961, 962, ..., 965. Considering the I_{REF1}
circuit 931 as a representative, it comprises the
transistor 961 as a current source for I_{REF1}. The I_{REF1}
level is obtained as a reproduction of the conduction
10   current of the reference cell 431. The reference cell
431 supplies a reference current I_{REF1} to a first leg 976
of the first current mirror that gets reproduced in a
second leg 977 thereof. The second leg 977 of the first
current mirror is interconnected with a first leg of the
15   second current mirror. Thus the reproduced reference
current is in turn reproduced in the second leg 941 of
the second mirror by the transistor 961. Generally, the
two current mirrors are of opposite polarity. For
example, when the REF1 cell 431 is an n-channel
20   transistor, the first current mirror comprises of two p-
channel transistors 981 and 982 of equal size "X", and
the second current mirror comprises of two n-channel
transistors 983 and 961 of equal size "W".

Figure 9G illustrates another embodiment in
25   which the different I_{REF} levels supplied by the second
current source of each branch are all generated from one
reference circuit 976. The reference circuit 976
provides a reference voltage that is applied to every
gate of the transistor 961, 962, ..., 965 of each branch
30   respectively. As in the embodiment illustrated in
Figure 9F, the reference voltage serves to turn on the
transistors. However, the different levels of I_{REF}'s
across the branches are now obtained by adjusting the
size of the transistors 961, 962, ..., 965. For
35   example, as illustrated in Figure 9G, the transistors
961, 962, ..., 965 respectively have sizes of I*W, J*W,

28

..., $K*W$, where $I{:}J{:}\ \ldots\ {:}K$ are respectively in the same ratios as $I_{REF1}{:}I_{REF2}{:}\ \ldots\ {:}I_{REFk}$. The single reference circuit 976 may be a constant voltage source or a circuit involving a reference cell similar to the REF

5   circuit 971 in Figure 9F. This applies under the normal current mirroring condition in which the transistors in each branch such as M81 and 961 are biased in the saturation region.

Figure 9H illustrates another embodiment in

10 which all the second current sources are the same across the branches but I_{CELL} is reproduced by the first current source into each branch with levels scaled according to the gradation of the reference current levels. The scaling is effected by adjusting the size of each second

15 transistor 911, 912, ..., 915. For example, as illustrated in Figure 9H, the second transistors 911, 912, ..., 915 respectively have sizes of $I*X$, $J*X$, ..., $K*X$, where X is the size of the first transistor 910 in the first leg 920 and $I{:}J{:}\ \ldots\ {:}K$ are respectively in

20 the same ratios as $I_{REF1}{:}I_{REF2}{:}\ \ldots\ {:}I_{REFk}$. Thus, only one REF circuit 976 is used across the branches, and furthermore, the sizes of all the transistors 961, 962, ..., 965 are now identical. The single reference circuit 976 may be a constant voltage source or may be

25 a circuit involving a reference cell similar to the REF circuit 971 in Figure 9F. In one implementation, the reference circuit 976 is such that each second current source 961, 962, ..., 965 is made to supply a current equal to the highest reference current level I_{REFk}. The

30 order of the outputs from the nodes is reversed relative to the embodiments illustrated in Figures 9D - 9G.

Figure 9I illustrates yet another embodiment of simultaneous multi-state sensing with a circuit similar to that in Figure 9G, except the identities of

35 the address memory cell and the IREF circuit are interchanged. In other words, in each branch, the

second current source such as 931, 932, ..., 935 now
supplies a reproduced I_{CELL}. This is achieved by means
of an addressed memory cell circuit 977 feeding a
reference voltage MC to every gate of the transistor
961, 962, ..., 965 of each branch respectively. The
circuit 977 is similar to the REF1 circuit 971 in Figure
9F, except the REF1 CELL 431 is now replaced by the
addressed memory cell 420. Similarly, the first current
source such as 911, 912, ..., 915 now supplies
respectively I_{REF1}, I_{REF2}, ..., I_{REFk}. The various I_{REF}'s are
obtained by a scaled reproduction of the current of an
IREF0 circuit 978. The scaling is effected by adjusting
the size of each second transistor 911, 912, ..., 915 in
the one-to-many current mirror. For example, as
illustrated in Figure 9I, the second transistors 911,
912, ..., 915 respectively have sizes of I*X, J*X, ...,
K*X, where X is the size of the first transistor 910 in
the first leg 920 and 1:I:J: ... :K are respectively in
the same ratios as $I_{REF0}:I_{REF1}:I_{REF2}: ... :I_{REFk}$. In general,
the IREF0 circuit 978 may be any current source which
supplies a current level of I_{REF0}. In one embodiment,
the IREF0 circuit is an EEprom cell programmable with a
reference current level, similar to that described in
connection with Figures 9A and 9B.

Another important feature of the present
invention serves to overcome the problems of endurance-
related stress. As explained previously, the erase,
program and read characteristics of each memory cell
depends on the cumulated stress endured over the number
of program/erase cycles the cell has been through. In
general, the memory cells are subjected to many more
program/erase cycles than the master reference cells.
The initially optimized reference levels will eventually
become misaligned to cause reading errors. The present
underlying inventive concept is to have the reference
levels also reflect the same cycling suffered by the

30

memory cells. This is achieved by the implementation of
local reference cells in addition to the master
reference cells. The local reference cells are
subjected to the same program/erase cycling as the
5   memory cells. Every time after an erase operation, the
reference levels in the master reference cells are re-
copied into the corresponding set of local reference
cells. Memory cells are then read with respect to the
reference levels of the closely tracking local reference
10   cells. In this way, the deviation in cell
characteristics after each program/erase cycle is
automatically compensated for. The proper partitioning
of the transforming threshold window is therefore
maintained so that the memory states can be read
15   correctly even after many cycles.

Figure 10 illustrates the local cells
referencing implementation for Flash EEprom. In the
Flash EEprom array 60 (Fig. 4), each group of memory
cells which is collectively erased or programmed is
20   called a sector. The term " Flash sector" is analogous
to the term "sector" used in magnetic disk storage
devices and they are used interchangeably here. The
EEprom array is grouped into Flash sectors such as 501,
503 and 505. While all memory cells in a Flash sector
25   suffer the same cycling, different Flash sectors may
undergo different cycling. In order to track each Flash
sector properly, a set of memory cells in each Flash
sector is set aside for use as local reference cells.
For example, after the Flash sector 503 has been erased,
30   the reference levels in the master reference cells 507
are re-programmed into the local reference cells
associated with the Flash sector 503. Until the next
erase cycle, the read circuits 513 will continue to read
the memory cells within the Flash sector 503 with
35   respect to the re-programmed reference levels.

31

Figures 11(1)-11(7) illustrates the algorithm
to re-program a sector's reference cells. In
particular, figures 11(1)-11(3) relate to erasing the
sector's local reference cells to their "erased states".
Thus in figure 11(1), a pulse of erasing voltage is
applied to all the sector's memory cells including the
local reference cells. In figure 11(2), all the local
reference cells are then read with respect to the master
references cells to verify if they have all been erased
to the "erased state". As long as one cell is found to
be otherwise, another pulse of erasing voltage will be
applied to all the cells. This process is repeated
until all the local reference cells in the sector are
verified to be in the "erased" state (figure 11(3)).

Figures 11(4)-11(7) relate to programming the
local reference cells in the sector. After all the
local reference cells in the sector have been verified
to be in the "erased" state, a pulse of programming
voltage is applied in figure 11(4) only to all the local
reference cells. This is followed in figure 11(5) by
reading the local reference cells with respect to the
master reference cells to verify if every one of the
local reference cells is programmed to the same state as
the corresponding master reference cell. For those
local reference cells not so verified, another pulse of
programming voltage is selectively applied to them alone
(figure 11(6)). This process is repeated until all the
local reference cells are correctly verified (figure
11(7)) to be programmed to the various breakpoint
threshold levels in the threshold window.

Once the local reference cells in the sector
have been re-programmed, they are used directly or
indirectly to erase verify, program verify or read the
sector's addressed memory cells.

Figure 12A illustrates one embodiment in which
the local reference cells are used directly to read or

32

program/erase verify the sector's memory cells. Thus,
during those operations, a parallel pair of switches 525
is enabled by a READ signal and the sense amplifier 440
will read the sector's addressed memory cells 523 with
5   respect to each of the sector's local reference cells
525.   During program/erase verify of the local
reference cells (as illustrated in figure 11), another
parallel pair of switches 527 enables reading of the
local reference cells 525 relative to the master
10 reference cells 529.

        Figure 12B illustrates the algorithm for using
the local reference cells directly to read or
program/erase verify the sector's addressed memory
cells.

15   Figure 13A illustrates an alternative
embodiment in which the local reference cells are used
indirectly to read the addressed memory cells. First
the master reference cells are erased and programmed
each to one of the desired multiple breakpoint
20 thresholds within the threshold window. Using these
master reference thresholds the local reference cells
within an erased sector of cells are each programmed to
one of the same desired multiple breakpoint thresholds.
Next the addressed cells in the sector are programmed
25 (written) with the desired data. The reading sequence
for the addressed cells in the sector then involves the
steps illustrated in Figure 13A.

        First each of the local reference cells 525 is
read relative to the corresponding master reference cell
30 531. This is effected by an enabling READ 1 signal to
a switch 533 connecting the local reference cells 525 to
the second leg 457 of the sense amplifier 440 with the
master reference 531 connected to the first leg 441 of
the sense amplifier. Auxiliary current source circuits
35 associated with each master reference cell are now used
to optimally bias the current through the first leg 441

33

of the sense amplifier to match the current in the
second leg 457. After the bias adjustment operation is
completed for all breakpoint threshold levels the
addressed cells in the sector are read relative to the
5   bias-adjusted master reference cells. This is effected
by disabling READ 1 to 533 and enabling READ signal to
switch 535. The advantage of this approach is that any
variations in V_{cc}, temperature, cycling fatigue or other
effects which may, over time, cause threshold deviations
10   between the master reference cells and the addressed
cells is eliminated prior to reading, since the local
reference cells (which track threshold deviations of the
addressed cells) are used to effectively readjust the
breakpoint thresholds of the master reference cells.
15   For example, this scheme permits programming of the
addressed cells when the master reference cells are
powered with $V_{cc}=5.5V$ and subsequently reading the
addressed cells with the master reference cells powered
at $V_{cc}=4.5V$. The difference of 1 volt in V_{cc}, which
20   would normally cause a change in the value of the
breakpoint thresholds, is neutralized by using the local
reference cells to bias adjust the master reference
cells to counteract this change at the time of reading.

Figures 13B and 13C show in more detail one
25   embodiment of the current biasing circuits such as 541,
543, 545 for the master reference cells 551, 553, 555.
Each biasing circuit acts as a current shunt for the
current in the master reference cell. For example, the
circuit 541 is tapped to the drain of the master
30   reference cell 551 through the line 561. It modifies the
current in line 562 to the sense amplifier (first leg)
either by sourcing current from V_{cc} or draining current
to V_{ss}. In the former case, the current in the line 562
is reduced, and otherwise for the latter case. As
35   biasing is being established for the master reference
551, any inequality in the currents in the two legs of

34

the sense amplifier can be communicated to outside the chip. This is detected by the controller (see figure 5) which in turn programs the biasing circuit 541 via the internal address bus 111 to subtract or add current in
5   the line 562 in order to equalize that of the local reference.

Figure 13C illustrates an embodiment of the biasing circuit such as the circuit 541. A bank of parallel transistors such as 571, 573, 575 are all
10   connected with their drains to V_{cc}, and their sources via switches such as 581, 583, 585 to the line 561. By selectively enabling the switches, different number of transistors may be used to subtract various amount of current from line 562. Similarly, another bank of
15   parallel transistors such as 591, 593, 595 are all connected with their sources to V_{ss}, and their drains via switches such as 601, 603, 605 to the line 561. By selectively enabling the switches, different number of transistors may be used to add various amount of current
20   to line 562. A decoder 609 is used to decode address from the internal address bus 111 to selectively enable the switches. The enabling signals are stored in latches 611, 613. In this way every time a sector is read, the master reference cells are re-biased relative
25   to the local reference cells, and used for reading the memory cells in the sector.

Figures 13D(1)-13D(4) illustrate the read algorithm for the alternative embodiment. The sector must previous had its local reference cells programmed
30   and verified relative to the master reference cells (figure 13D(1)). Accordingly, each of the master reference cells is then read relative to the local reference cells (figure 13D(2)). The master reference cells are biased to equalize the current to that of the
35   corresponding local reference cells (figure 13D(3)). Subsequently, the memory cells in the sector are read

35

relative to the biased master reference cells(figure 13D(4)).

The read circuits and operation described are also employed in the programming and erasing of the memory cells, particularly in the verifying part of the operation. As described previously, programming is performed in small steps, with reading of the state programmed in between to verify if the desired state has been reached. As soon as the programmed state is verified correctly, programming stops. Similarly, erasing is performed in small steps, with reading of the state of erase in between to verify if the "erased" state has been reach. Once the "erased" state is verified correctly, erasing stops.

As described previously, only K-1 breakpoint threshold levels are required to partition the threshold window into K regions, thereby allowing the memory cell to store K states. According to one aspect of the present invention, however, in the multi-state case where the threshold window is more finely partitioned, it is preferable to use K threshold levels for K state. The extra threshold level is used to distinguish the "erased" state from the state with the lowest threshold level. This prevents over-erasing and thus over-stressing the cell since erasing will stop once the "erased" state is reached. The selective inhibition of individual cells for erase does not apply to the Flash EEprom case where at least a sector must be erased each time. It is suitable those EEprom arrays where the memory cells can be individually addressed for erase.

According to another feature of the invention, after a memory cell has been erased to the "erased" state, it is programmed slightly to bring the cell to the state with the lowest threshold level (ground state) adjacent the "erased" state. This has two advantages. First, the threshold levels of the ground state of all

36

the memory cells, being confined between the same two breakpoint threshold levels, are well-defined and not widely scattered.   This provide an uniform starting point for subsequent programming of the cells.
5   Secondly, all cells get some programming, thereby preventing those cells which tend to have the ground state stored in them, for example, from losing track with the rest with regard to program/erase cycling and endurance history.
10

On Chip Program Verify

As mentioned before, programming of an EEprom cell to a desired state is preferably performed in small steps starting from the "erase" state.   After each
15   programming step, the cell under programming is read to verify if the desired state has been reached. If it has not, further programming and verifying will be repeated until it is so verified.

Referring to the system diagram illustrated in
20   figure 5, the EEprom chip 130 is under the control of the controller 140.   They are linked serially by the serial in line 251 and serial out line 253.   In prior art EEprom devices, after each programming step, the state attained in the cell under programming is read and
25   sent back to the controller 140 or the CPU 160 for verification with the desired state. This scheme places a heavy penalty on speed especially in view of the serial link.

In   the   present   invention,   the   program
30   verification is optimized by programming a chunk (typically several bytes) of cells in parallel followed by verifying in parallel and on chip. The parallel programming is implemented by a selective programming circuit which disables programming of those cells in the
35   chunk whose states have already been verified correctly. This   feature   is   essential   in   a   multi-state

37

implementation, because some cells will reach their desired state earlier than others, and will continue pass the desired state if not stopped. After the whole chunk of cells have been verified correctly, logic on
5   chip communicates this fact to the controller, whereby programming of the next chunk of cells may commence. In this way, in between each programming step data does not need to be shuttled between the EEprom chip and the controller, and program verification speed is greatly
10 enhanced.

Figure 14 illustrates the program and verify paths for a chunk of n cells in parallel. The same numerals are used for corresponding modules in the system diagram of figure 5. The EEprom array 60 is
15 addressed by N cells at a time. For example, N may be 64 cells wide. In a 512 bytes Flash sector, consisting of 4 rows of 1024 cells, there will be 64 chunks of 64 cells. The source multiplexer 107 selectively connects the N sources of one addressed chunk of cells to the
20 source voltage V_S in line 103. Similarly, the drain multiplexer 109 selectively makes the N drains of the chunk accessible through an N-channel data path 105. The data path 105 is accessed by the program circuit with inhibit 210 during programming and by read circuits
25 220 during reading, program verifying or erase verifying.

Referring again to the system diagram in figure 5, programming is under the control of the controller 140. The data to be programmed into the
30 sector is sent chunk by chunk. The controller first sends a first chunk of N*L serial data bits together with addresses, control and timing information to the EEprom chip 130. L is the number of binary bits encoded per memory cell. For example, L=1 for a 2-state cell,
35 and L=2 for a 4-state cell. Thus if N=64 and L=2, the chunk of data bits will be 128 bits wide. The N*L data

38

bits are stored in latches and shift registers 190 where the serial bits are converted to N*L parallel bits. These data will be required for program verify in conjunction with the read circuits 220, bit decoder 230,

5   compare circuit 200 and the program circuit with inhibit 210.

        The program algorithm for a chunk of N cells is best described by referring to both the system diagram of figure 5 and figures 15(1)-15(7) which

10   illustrate the algorithm itself. As mentioned in an earlier section, prior to programming the sector, the whole sector must be erased and all cells in it verified to be in the "erased" state (figure 15(1)). This is followed in figure 15(2) by programming the sector local

15   reference cells (as shown in figures 11(1)-(3)). In figure 15(3), the N*L bits of parallel data is latched in latches 190. In figure 15(4), the read circuits 220 access the N-channel data path 105 to read the states in the N chunk of cells. The read algorithm has already

20   been described in conjunction with figure 12B or figure 13D. The N-cell reads generates N*K (K=number of states per cell) output states. These are decoded by bit decoder 230 into N*L binary bits. In figure 15(5), the N*L read bits are compared bit by bit with the N*L

25   program data bits from latches 190 by compare circuit 200. In figure 15(6), if any read bit fails to compare with the program data bit, a further programming voltage pulse from the program circuit 210 is applied simultaneously to the chunk of cells. However, an

30   inhibit circuit within the program circuit 210 selectively blocks programming to those cells whose bits are correctly verified with the programmed data bits. Thus, only the unverified cells are programmed each time. Programming and verification are repeated until

35   all the cells are correctly verified in figure 15(7).

39

Figure 16 shows one embodiment of the compare circuit 200 of figure 5 in more detail. The circuit 200 comprises N cell compare modules such as 701, 703, one for each of the N cells in the chunk. In each cell
5   compare module such as the module 701, the L read bits (L=number of binary bits encoded for each cell) are compared bit by bit with the corresponding program data bits. This is performed by L XOR gates such as 711, 713, 715. The output of these XOR gates pass through an
10 NOR gate 717 such that a "1" appears at the output of NOR gate 717 whenever all the L bits are verified, and a "0" appears when otherwise. When the control signal VERIFY is true, this result is latched to a latch 721 such that the same result at the output of NOR gate 717
15 is available at the cell compare module's output 725. The compare circuit 200 performs the comparisons of L bits in parallel. The N compare module's outputs such as 725, 727 are available at an N-channel output line 731 to be fed to the program circuit with inhibit 210 of
20 figure 5.

At the same time, the N outputs such as 725, 727 are passed through an AND gate 733 so that its single output 735 results in a "1" when all N cells are verified and a "0" when otherwise. Referring also to
25 figure 5, the single output 735 is used to signal the controller 140 that all N cells in the chunk of data have been correctly verified. The signal in output 735 is sent through the serial out line 253 via AND gate 240 during a VERIFY operation.

30 At power-up or at the end of program/verify of a chunk of data, all cell compare module's outputs such as 725, 727 are reset to the "not-verified" state of "0". This is achieved by pulling the node 726 to V_{ss} (0V) by means of the RESET signal in line 727 to a
35 transistor 729.

40

Figure 17 shows one embodiment of the program circuit with inhibit 210 of figure 5 in more detail. The program circuit 210 comprises N program with inhibit modules such as 801, 803. As illustrated in Table 1 and 2, in order to program the N cells, a voltage V_{PD} must be applied to each of the N cells' drain and a voltage V_{PG} applied to the control gates. Each program module such as 801 serves to selectively pass V_{PD} on a line 805 to one of the drains through the one of the N-channel data path 105. Since V_{PD} is typically about 8V to 9V which is higher than V_{CC}, the latter cannot be used to turn on the transistor switch 807. Rather the higher voltage V_{CG} (about 12V) is used to enable switch 807. V_{CG} in line 801 is itself enabled by an AND gate when both the program control signal PGM in line 813 is true and the signal in line 731 is a "0". Since the signal in line 731 is from the output of the cell compare module 701 shown in figure 16, it follows that V_{PD} will be selectively passed onto those cells which are not yet verified. In this way, every time a programming pulse is applied, it is only applied to those cells which have not yet reached their intended states. This selective programming feature is especially necessary in implementing parallel programming and on chip verification in the multi-state case.

Variable Control of Voltage to the Control Gate

The system diagram of figure 5 in conjunction with Tables 1 and 2 illustrate how various voltages are applied to the EEprom array 60 to perform the basic functions of the EEprom. Prior art EEprom devices only allow the voltage supplied to the control gate V_{CG} to assume one of two voltages, namely V_{CC} or the higher programming voltage of about 12V.

In another aspect of the present invention, the voltage supplied to the control gate V_{CG} is allowing

41

to be independently and continuously variable over a
wide range of voltages. This is provided by V_{PG} from the
controller 140. In particular V_{CG} in a line 83 is fed
from V_{PG} which is in turn supplied by the controller from
5   a line 901. Table 2 shows V_{PG} to assume various voltages
under different functions of the EEprom.

The variability of V_{CG} is particularly
advantageous in program and erase margining schemes. In
program margining, the read during program verify is
10   done with V_{CG} at a slightly higher voltage than the
standard V_{CC}. This helps to place the programmed
threshold well into the state by programming past the
breakpoint threshold level with a slight margin. In
erase verify, the cell is verified with a somewhat
15   reduced V_{CG} to put the cell well into the "erased" state.
Furthermore, margining can be used to offset the charge
retention problem described earlier (Figure 8B).

As mentioned before, prior art EEproms
typically employ V_{CC} to feed V_{CG} during program or erase
20   verify. In order to do margining, V_{CC} itself needs to
be ramped up or reduced. This practice produces
inaccurate results in the reading circuits since they
are also driven by V_{CC}.

In the present invention, the variability of
25   V_{CG} independent of voltages supplied to the reading
circuit produce more accurate and reliable results.

Furthermore, the wide range of V_{CG} is useful
during testing and diagnostic of the EEprom. It allows
the full range of the programmed cell's threshold to be
30   measured easily by continuing to increase V_{CG} (up to the
maximum limited by the device's junction breakdown).

While the embodiments of this invention that
have been described are the preferred implementations,
those skilled in the art will understand that variations
35   thereof may also be possible. Therefore, the invention

42

is entitled to protection within the full scope of the appended claims.

43

WHAT IS CLAIMED IS:

1.   In an array of a plurality of addressable
semiconductor electrically erasable and programmable
memory cells of the type having a source, a drain, a
control gate, a floating gate capable of retaining a
5   charge level programmed into it during use of the
memory, and an erase electrode capable of removing
charge from said floating gate, a system for reading
whether the programmed state of an addressed cell is
greater than or less than a predetermined threshold,
10   comprising:
        a reference memory cell,
        means responsive to erasure and reprogramming
of said memory cell array for erasing and reprogramming
said reference cell with a charge that is substantially
15   equal to or proportional to said threshold, and
        reading means responsive to said reference
cell for comparing the charge level of an addressed cell
with that of said reference cell.

2.   The system as in claim 1, wherein the
reading system is also part of a system for programming
the memory cells in which each addressed cell is
programmed to a desired state by repetitive sequence of
5   programming in a small step and reading to verify the
state programmed until the desired state is reached.

3.   The system as in claim 1, wherein the
reading system is also part of an erasing system in
which each addressed cell is erased to the erased state
by repetitive sequence of erasing in a small step and
5   reading to verify the state erased until the erased
state is reached.

44

4.   The memory cell array reading system according to claim 1, wherein the charge level stored in said reference cell is electrically erasable and programmable from outside of said memory.

5.   The memory cell array reading system according to claim 1, which additionally comprises at least one master reference cell that is erasable and programmable from outside the memory with a charge that is substantially equal to or proportional to said threshold, and wherein said reference cell reprogramming means includes means for programming said reference cell from said master reference cell.

6.   The memory cell array reading system according to claim 5, wherein said reading means includes means for directly comparing the charge level of an addressed cell with that of said reference cell.

7.   The memory cell array reading system according to claim 5, wherein said reading means includes means for adjusting the predetermined threshold of said master reference cell to substantially match that of said reference cell, and means for comparing the addressed cell to the adjusted threshold of said master reference cell.

8.   In an array of a plurality of addressable semiconductor electrically erasable and programmable memory cells of the type having a source, a drain, a control gate, a floating gate capable of retaining a charge level programmed into it during use of the memory, and an erase electrode capable of removing charge from said floating gate, a system for reading the stored charge of an addressed cell within ranges defined

45

by at least two predetermined threshold levels,
10   comprising:

        at least two reference memory cells that are
each respectively programmed with a charge that is
substantially equal to or proportional to each of said
at least two thresholds; and

15   means responsive to said at least two
reference cells for comparing the charge level of an
addressed cell with that of each of said reference
cells, thereby to determine which of said plurality of
said stored ranges that the addressed cell lies, whereby
20   more than a single bit of data is stored and read from
each of the addressed cell.

        9.   The system as in claim 8, wherein the
reading system is also part of a system for programming
the memory cells in which each addressed cell is
programmed to a desired state by repetitive sequence of
5   programming in a small step and reading to verify the
state programmed until the desired state is reached.

        10.   The system as in claim 8, wherein the
reading system is also part of an erasing system in
which each addressed cell is erased to the erased state
by repetitive sequence of erasing in a small step and
5   reading to verify the state erased until the erased
state is reached.

        11.   The memory cell reading system according
to claim 8, wherein said comparing means includes means
for comparing said addressed cell with the charge levels
of said at least two reference memory cells one at a
5   time in sequence.

        12.   The memory cell reading system according
to claim 8, which additionally includes means for

46

independently erasing and programming said at least two reference memory cells to said at least two thresholds
5 from outside of said memory.

13.   In an array of a plurality of addressable semiconductor electrically erasable and programmable memory cells of the type having a source, a drain, a control gate, a floating gate capable of retaining a
5 charge level programmed into it during use of the memory, and an erase electrode capable of removing charge from said floating gate, a system for reading the stored charge of an addressed cell within two ranges defined   by   one   predetermined   threshold   level,
10 comprising:
    at least one reference memory cell that is programmed with a charge that is substantially equal to or proportional to said threshold; and
    means responsive to said reference cell for
15 comparing the charge level of an addressed cell with that of said reference cell, thereby to determine which of said two stored ranges that the stored charge of the addressed cell lies, whereby a single bit of data is stored and read from each of the addressed cell.

14.   The system as in claim 13, wherein the reading system is also part of a system for programming the memory cells in which each addressed cell is programmed to a desired state by repetitive sequence of
5 programming in a small step and reading to verify the state programmed until the desired state is reached.

15.   The system as in claim 13, wherein the reading system is also part of an erasing system in which each addressed cell is erased to the erased state by repetitive sequence of erasing in a small step and

47

5 reading to verify the state erased until the erased
state is reached.

16. In an array of a plurality of addressable
semiconductor electrically erasable and programmable
memory cells of the type having a source, a drain, a
conductance between the source and drain that is
5 controlled by the level of charge programmed onto a
floating gate, and having a control gate and an erase
electrode, a system for reading the state of an
addressed cell by measuring the level of current passing
therethrough, comprising:
10 means for passing current between the source
and drain of the addressed cell in a manner to provide
a current level that is proportional to the level of
charge upon the floating gate of the addressed cell,
at least two of said memory cells being
15 provided as reference cells with charges programmed on
their respective floating gates corresponding to
respective at least two predetermined threshold levels,
and
means connected to both of the addressed and
20 reference cells for comparing the current flowing
through the addressed cell with that flowing through
each of said reference cells whereby the programmed
charge of said addressed cell is determined to lie
within one of at least three levels defined by said
25 thresholds, thereby to store and read at least one and
a half bits of information from the addressed cell.

17. The system as in claim 16, wherein the
reading system is also part of a system for programming
the memory cells in which each addressed cell is
programmed to a desired state by repetitive sequence of
5 programming in a small step and reading to verify the
state programmed until the desired state is reached.

48

18. The system as in claim 16, wherein the reading system is also part of an erasing system in which each addressed cell is erased to the erased state by repetitive sequence of erasing in a small step and
5   reading to verify the state erased until the erased state is reached.

19. The memory array cell reading system according to claim 16, wherein said comparing means includes a current mirror circuit connecting said addressed cell and said reference cells.

20. An EEprom memory system on an integrated circuit chip, comprising:
       a plurality of groups of individually addressable EEprom memory cells,
5   one or more EEprom group reference cells provided as part of each of said group of memory cells,
       means responsive to signals from outside of said chip for programming said individually addressable cells to one of at least two conduction states,
10   means responsive to signals from outside of said chip for simultaneously erasing all the addressable and reference cells of a designated group,
       one or more master EEprom reference cells,
       means responsive to signals from outside of
15   said chip for erasing and programming different threshold levels on each of said one or more master reference cells that correspond to breakpoints between said at least three conductive states,
       means responsive to said individually
20   addressable cells of a group being programmed for programming that group's reference cells to the levels of said master reference cells,
       means responsive to signals from outside of said chip for reading an addressed individually

49

25   addressable cell of a given group of cells by comparison
with the reference cells of said given group.

21.   The system as in claim 20, wherein the
reading means is also part of a system for programming
the memory cells in which each addressed cell is
programmed to a desired state by repetitive sequence of
5   programming in a small step and reading to verify the
state programmed until the desired state is reached.

22.   The system as in claim 20, wherein the
reading means is also part of an erasing system in which
each addressed cell is erased to the erased state by
repetitive sequence of erasing in a small step and
5   reading to verify the state erased until the erased
state is reached.

23.   The memory system according to claim 20
wherein said reading means includes means for directly
comparing an addressed cell with the reference cells of
the given group.

24.   The memory system according to claim 20
wherein said reading means includes means for adjusting
the master reference cells to correspond to the group
reference cells, and means for comparing an addressed
5   cell with the so adjusted master reference cells.

25.   The system as in claim 24, wherein the
reading means is also part of a system for programming
the memory cells in which each addressed cell is
programmed to a desired state by repetitive sequence of
5   programming in a small step and reading to verify the
state programmed until the desired state is reached.

50

26.   The system as in claim 24, wherein the reading means is also part of an erasing system in which each addressed cell is erased to the erased state by repetitive sequence of erasing in a small step and
5   reading to verify the state erased until the erased state is reached.

27.   In an array of addressable semiconductor electrically erasable and programmable memory (EEprom) cells on an integrated circuit chip, the memory cell being of the type having a source, a drain, a control
5   gate and an erase electrode receptive to specific voltage conditions for reading, programming and erasing of data in the cell, and having a floating gate capable of retaining a specific charge level corresponding to a specific memory state of the cell, such that a specific
10   memory state is achieved by increment or decrement of the charge level with successive applications of programming or erasing voltage conditions, a system for programming data to EEprom cells comprising:
          means for temporarily storing a chunk of data
15   for programming a plurality of addressed cells;
          means for programming in parallel the stored chunk of data into the plurality of addressed cells;
          means for verifying the programmed data in each of the plurality of addressed cells with the chunk
20   of stored data;
          means for enabling further programming and verifying in parallel to one or more of the addressed cells until all the plurality of addressed cells are verified.

28.   The system for programming the EEprom cells as in claim 27, wherein the system resides on the EEprom integrated circuit chip.

51

29. The system for programming the EEprom cells as in claim 28, wherein the verifying means includes a system for reading the stored charge of an addressed cell within ranges defined by one or more predetermined threshold levels, comprising:

one or more reference memory cells that are each respectively programmed with a charge that is substantially equal to or proportional to each of said one or more thresholds; and

means responsive to said one or more reference cells for comparing the charge level of an addressed cell with that of each of said reference cells, thereby to determine which of said plurality of said stored ranges that the addressed cell lies.

30. The system for programming the EEprom cells as in claim 28, wherein the memory cells are grouped such that all cells in the group are erasable together and wherein the verifying means includes a reading circuit comprising:

one or more EEprom group reference cells provided as part of each of said group of memory cells,

means responsive to signals from outside of said chip for programming said individually addressable cells to one of at least two conduction states,

means responsive to signals from outside of said chip for simultaneously erasing all the addressable and reference cells of a designated group,

one or more master EEprom reference cells,

means responsive to signals from outside of said chip for erasing and programming different threshold levels on each of said one or more master reference cells that correspond to breakpoints between said at least two conduction states,

means responsive to said individually addressable cells of a group being programmed for

52

programming that group's reference cells to the levels
of said master reference cells,

25   means for reading an addressed individually
addressable cell of a given group of cells by comparison
with the reference cells of said given group.

          31.   The memory system according to claim 30
wherein said reading means includes means for adjusting
the master reference cells to correspond to the group
reference cells, and means for comparing an addressed
5   cell with the so adjusted master reference cells.

          32.   The system for programming the EEprom
cells as in claim 28, wherein the cells have binary
states, and means for enabling further programming
applies to all addressed cells in parallel until all the
5   plurality of addressed cells are verified.

          33.   The system for programming the EEprom
cells as in claim 28, further comprising means on chip
for individually inhibiting programming of any addressed
cell   already   verified,   while   enabling   further
5   programming in parallel to all other addressed cells not
yet verified.

          34.   The system according to claim 33, wherein
the memory cells have more than two states.

          35.   The system according to claim 33, wherein
the memory cells have binary states.

          36.   In an array of addressable semiconductor
electrically erasable and programmable memory cells on
an integrated circuit chip, the memory cell being of the
type having a source, a drain, a control gate and an
5   erase electrode receptive to specific voltage conditions

53

for reading, programming and erasing of data in the
cell, and having a floating gate capable of retaining a
specific charge level corresponding to a specific memory
state of the cell, such that a specific memory state is
10   achieved by increment or decrement of the charge level
with successive applications of programming or erasing
voltage conditions, a system for erasing the EEprom
memory cells comprising:

        means for erasing in parallel a plurality of
15   addressed cells;

        means for verifying the memory state in each
of the plurality of addressed cells;

        means for enabling further erasing in parallel
to one or more of the addressed cells until all the
20   plurality of addressed cells are verified to be in an
erased state.

        37.   The system for erasing the EEprom cells
as in claim 36, wherein the system resides on the EEprom
integrated circuit chip.

        38.   The system for erasing the EEprom cells
as in claim 36, wherein the system resides outside the
EEprom integrated circuit chip.

        39.   The system for erasing the EEprom cells
as in claim 38, wherein the verifying means includes a
system for reading the stored charge of an addressed
cell within ranges defined by one or more predetermined
5   threshold levels, comprising:

        one or more reference memory cells that are
each respectively programmed with a charge that is
substantially equal to or proportional to each of said
one or more thresholds; and
10   means responsive to said one or more reference
cells for comparing the charge level of an addressed

54

cell with that of each of said reference cells, thereby to determine which of said plurality of said stored ranges that the addressed cell lies.

40.   The system for erasing the EEprom cells as in claim 38, wherein the memory cells are grouped such that all cells in the group are erasable together and wherein the verifying means includes a reading
5   circuit comprising:

one or more EEprom group reference cells provided as part of each of said group of memory cells,

means responsive to signals from outside of said chip for programming said individually addressable
10   cells to one of at least two conduction states,

means responsive to signals from outside of said chip for simultaneously erasing all the addressable and reference cells of a designated group,

one or more master EEprom reference cells,
15   means responsive to signals from outside of said chip for erasing and programming different threshold levels on each of said one or more master reference cells that correspond to breakpoints between said at least two conduction states,
20   means responsive to said individually addressable cells of a group being programmed for programming that group's reference cells to the levels of said master reference cells,

means for reading an addressed individually
25   addressable cell of a given group of cells by comparison with the reference cells of said given group.

41.   The memory system according to claim 40 wherein said reading means includes means for adjusting the master reference cells to correspond to the group reference cells, and means for comparing an addressed
5   cell with the so adjusted master reference cells.

55

42. The system for erasing the EEprom cells
as in claim 38, wherein the cells have binary states,
and means for enabling further erasing applies to all
addressed cells in parallel until all the plurality of
5   addressed cells are verified.

43. The system for programming the EEprom
cells as in claim 38, further comprising means on chip
for individually inhibiting erasing of any addressed
cell already verified, while enabling further erasing in
5   parallel to all other addressed cells not yet
verified.

44. The system according to claim 43, wherein
the memory cells have more than two states.

45. The system according to claim 43, wherein
the memory cells have binary states.

46. The system for erasing the EEprom cells
as in claim 36, further comprising means for programming
the cells in the erased state to the memory state
adjacent the erased state, thereby ensuring uniformity
5   of threshold level in each of the erased cells and that
each cell is subject to similar amount of program/erase
stress.

47. In a EEprom system including an array of
addressable semiconductor electrically erasable and
programmable memory cells on an integrated circuit chip,
wherein the improvement in programming a plurality of
5   addressed memory cells comprises:
a controller for controlling the operation of
the memory cells;
means for temporarily storing on chip a chunk
of data serially transferred from the controller;

56

10   means for programming in parallel the stored
chunk of data into the plurality of addressed cells;
        means for verifying on chip the programmed
data in each of the plurality of addressed cells with
the chunk of stored data;
15   means for enabling further programming in
parallel to one or more of the addressed cells until all
the plurality of addressed cells are verified; and
        means for outputting a signal from the chip to
the controller to indicate that all the plurality of
20 addressed cells are verified.

        48. The system according to claim 47 wherein
the cells have binary states, and means for enabling
further programming applies to all addressed cells in
parallel until all the plurality of addressed cells are
5   verified.

        49. The system according to claim 47 wherein
the memory cells have more than two states, thereby
enabling the storing of more than one bit per memory
cell.

        50. The EEprom system according to claim 47
further comprising means on chip for individually
inhibiting programming of any addressed cell already
verified, while enabling further programming in parallel
5   of all other addressed cells not yet verified.

        51. In a EEprom system including an array of
addressable semiconductor electrically erasable and
programmable memory cells on an integrated circuit chip,
wherein the improvement in erasing a plurality of
5   addressed memory cells comprises:
        a controller for controlling the operation of
the memory cells;

57

means for erasing in parallel the plurality of
addressed cells;

10   means for verifying on chip the state in each
of the plurality of addressed cells;

means for enabling further erasing in parallel
to one or more of the addressed cells until all the
plurality of addressed cells are verified to be in the
15   erased state.

52. An improved method for programming erased
memory cells in an EEprom array integrated circuit chip
comprising the steps of:

temporarily storing on chip a chunk of data
5   for programming a plurality of addressed memory cells;

reading in parallel the content of the
plurality of addressed memory cells;

verifying on chip by comparison in parallel of
the read content with the corresponding stored chunk of
10   data;

selectively applying in parallel a pulse of
programming voltage only to the addressed cells which do
not verify; and

repeating   the   verifying   and   selective
15   programming steps until all the plurality of addressed
cells are verified.

53. An improved method for programming erased
memory cells in an EEprom array integrated circuit chip
comprising the steps of:

temporarily storing on chip a chunk of data
5   for programming a plurality of addressed memory cells;

applying in parallel a pulse of programming
voltage to all the addressed cells;

reading in parallel the content of the
plurality of addressed memory cells;

58

verifying on chip by comparison in parallel of the read content with the corresponding stored chunk of data;

selectively applying in parallel a pulse of programming voltage only to the addressed cells which do not verify; and

repeating the verifying and selective programming steps until all the plurality of addressed cells are verified.

54. In a EEprom system including an array of addressable semiconductor electrically erasable and programmable memory cells on an integrated circuit chip and a controller for controlling the operation and voltage conditions of the memory cells, an improved method for programming erased memory cells comprising the steps of:

transferring serially from the controller to the chip a chunk of data for programming a plurality of addressed memory;

storing on chip the chunk of data;

applying in parallel a pulse of programming voltage to all the addressed cells;

reading in parallel the content of the plurality of addressed memory cells;

verifying on chip by comparison in parallel of the read content with the corresponding stored chunk of data;

selectively applying in parallel a pulse of programming voltage only to the addressed cells which do not verify;

repeating the verifying and selective programming steps until all the plurality of addressed cells are verified; and

59

25   outputting a signal from the chip to indicate
that all the plurality of the addressed cells are
verified.

55. The system according to claim 27 wherein
the voltage supplied to the control gate is variable and
independent of other voltages.

56. The system according to claim 55 wherein
the voltage supplied to the control gate during
verification of programmed data is sufficiently greater
than that supplied during normal read such that a
5   programmed cell is not just marginally in the programmed
state.

57. The system according to claim 36 wherein
the voltage supplied to the control gate is variable and
independent of other voltages.

58. The system according to claim 57 wherein
the voltage supplied to the control gate during
verification of erased cells is sufficiently less than
that supplied during normal read such that an erased
5   cell is not just marginally in the erased state.

59. A circuit for sensing a test current
relative to a plurality of predetermined current levels,
comprising:
          a one-to-many current mirror means for
5   reproducing a test current into one or more reproduced
currents, said current mirror having a first leg for
carrying the test current and a second leg comprising a
plurality of branches, such that each branch is
associated with a reference current level;
10   a first current source at each branch for
reproducing a reproduced current therein, said

60

reproduced current being substantially similar to the
test current in the first leg;

15   a second current source at each branch for
providing a reference current having one of the
predetermined reference current levels; and

            means for simultaneously detecting in each
branch a relatively high or low voltage at a node
between first and second current sources, the relatively
20   high or low voltage corresponding to whether the
reproduced current similar to the test current provided
by the first current source has magnitude greater or
less than that of the reference current provided by the
second current source.

        60.   A circuit as in claim 59, wherein said
test current is provided by a conduction current of a
memory cell, said conduction current designating a
memory state thereof.

        61.   A circuit as in claim 60, wherein said
memory cell is an EEprom or a Flash EEprom.

        62.   A circuit as in claim 59, wherein said
second current source in each branch is a programmable
reference current source.

        63.   A circuit as in claim 62, wherein said
test current is provided by a conduction current of a
memory cell, said conduction current designating a
memory state thereof.

        64.   A circuit as in claim 5, wherein said
memory cell is an EEprom or a Flash EEprom.

        65.   A circuit as in claim 62, wherein said
programmable reference current source is provided by a

61

conduction current of a memory cell, said conduction
current having conduction current designating a memory
5   state thereof.

66.   A circuit as in claim 65, wherein said
memory cell is an EEprom or a Flash EEprom.

67.   A circuit as in claim 62, wherein said
programmable reference current source is duplicated from
an original programmable reference current source by a
circuit which comprises:
5   first and second one-to-one current mirror
means each having only one branch for each of its two
legs, said first and second one-to-one current means
being interconnected by one of their two legs, and said
second current source being provided by connection to
10   the other leg of the second one-to-one current mirror
means; and
         an original programmable reference current
source for providing a reference current, said original
programmable reference current source being connected to
15   the other leg of the first one-to-one current mirror
means, thereby mirroring substantially similar reference
current into the interconnected leg, and in turn
mirroring same into the other leg of the second one-to-
one current means, thereby providing said second current
20   source with substantially similar current as that of the
original programmable reference current source.

68.   A circuit as in claim 67, wherein said
test current is provided by a conduction current of a
memory cell designating a memory state thereof.

69.   A circuit as in claim 68, wherein said
memory cell is an EEprom or a Flash EEprom.

62

70.   A circuit as in claim 67, wherein said programmable reference current source is provided by a memory cell having conduction current designating a memory state thereof.

71.   A circuit as in claim 70, wherein said memory cell is an EEprom or a Flash EEprom.

72.   A circuit as in claim 67, wherein the one-to-many current mirror means and the first one-to-one current mirror means comprise transistors having the same polarity, and the second one-to-one current mirror means comprise transistors having the opposite polarity.

73.   A circuit as in claim 59, wherein said second current source for providing a predetermined reference current in each branch comprises:

        a transistor of a predetermined size having a source, a drain and a gate; and

        means for applying a predetermined reference voltage to the gate for producing one of the predetermined reference currents through the source and drain of the transistor of predetermined size;

        said predetermined reference voltage being constant across the branches and said predetermined size differing across the branches so as to produce the plurality of predetermined reference currents among the branches.

74.   A circuit as in claim 73, wherein said test current is provided by a conduction current of a memory cell, said conduction current designating a memory state thereof.

75.   A circuit as in claim 74, wherein said memory cell is an EEprom or a Flash EEprom.

63

76. A circuit as in claim 73, wherein said constant reference current is provided by a conduction current of a memory cell, said conduction current designating a memory state thereof.

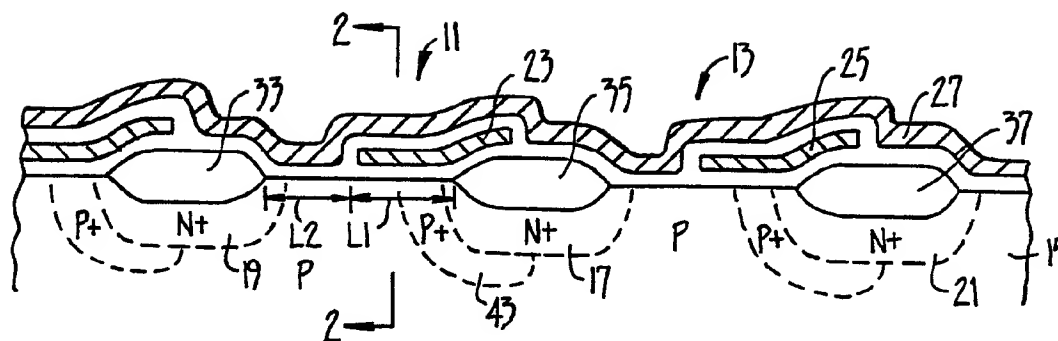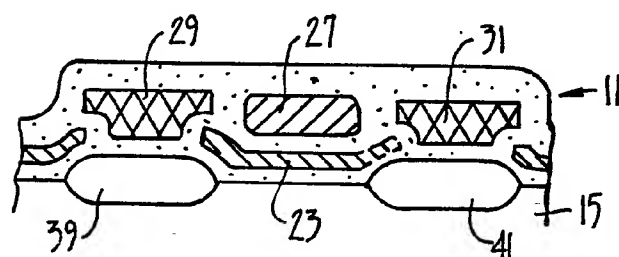77. A circuit as in claim 76, wherein said memory cell is an EEprom or a Flash EEprom.

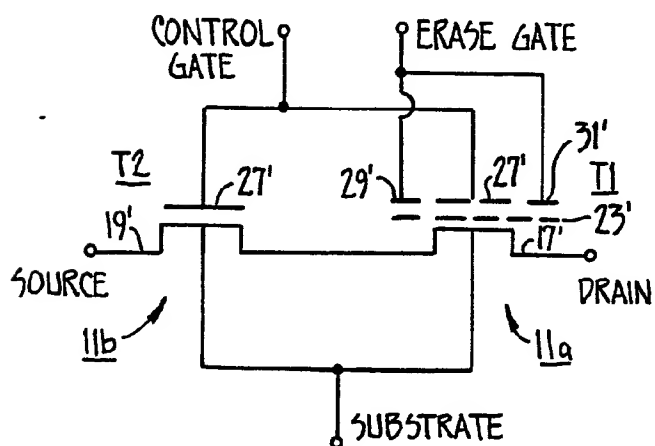78. A circuit for sensing a test current relative to a plurality of predetermined reference current levels, each of the plurality of reference current levels being scalable by a multiplication factor
5   from that of a lowest reference current level thereof, comprising:
        a one-to-many current mirror means for reproducing a test current into one or more reproduced currents, said current mirror having a first leg for
10   carrying the test current and a second leg comprising a plurality of branches, such that each branch is associated with a reference current level and a multiplication factor;
        a first current source at each branch for
15   reproducing a reproduced current therein, said reproduced current being scalable from the test current in the first leg by the associated multiplication factor;
        a second current source at each branch for
20   providing a highest reference current level from among said plurality of reference current levels; and
        means for simultaneously detecting in each branch a relatively high or low voltage at a node between first and second current sources, the relatively
25   high or low voltage corresponding to whether the reproduced current scaled from the test current provided by the first current source has magnitude greater or

64

less than that of the highest reference current provided by the second current source.

79. A circuit as in claim 78, wherein said test current is provided by a conduction current of a memory cell, said conduction current designating a memory state thereof.

80. A circuit as in claim 79, wherein said memory cell is an EEprom or a Flash EEprom.

81. A circuit as in claim 78, wherein said highest reference current is provided by a conduction current of a memory cell, said conduction current designating a memory state thereof.

82. A circuit as in claim 81, wherein said memory cell is an EEprom or a Flash EEprom.

83. A circuit for sensing a test current relative to a plurality of predetermined reference current levels, each of the plurality of reference current levels being scalable by a multiplication factor
5   from that of a given reference current level thereof, comprising:
         a one-to-many current mirror means for reproducing the given reference current level into one or more reproduced currents, said current mirror having
10   a first leg for carrying the given reference current and a second leg comprising a plurality of branches, such that each branch is associated with a reference current level and a multiplication factor;
         a first current source at each branch for
15   reproducing a reproduced current therein, said reproduced current being scalable from the given

65

reference current level in the first leg by the associated multiplication factor;

20   a second current source at each branch for providing the test current; and

        means for simultaneously detecting in each branch a relatively high or low voltage at a node between first and second current sources, the relatively high or low voltage corresponding to whether the

25   reproduced current scaled from the lowest reference current level provided by the first current source has magnitude greater or less than that of the test current provided by the second current source.

        84.   A circuit as in claim 83, wherein said test current is provided by a conduction current of a memory cell, said conduction current designating a memory state thereof.

        85.   A circuit as in claim 84, wherein said memory cell is an EEprom or a Flash EEprom.

        86.   A circuit as in claim 83, wherein said lowest reference current is provided by a conduction current of a memory cell, said conduction current designating a memory state thereof.

        87.   A circuit as in claim 86, wherein said memory cell is an EEprom or a Flash EEprom.

FIG._1.

FIG._2.

FIG._3.
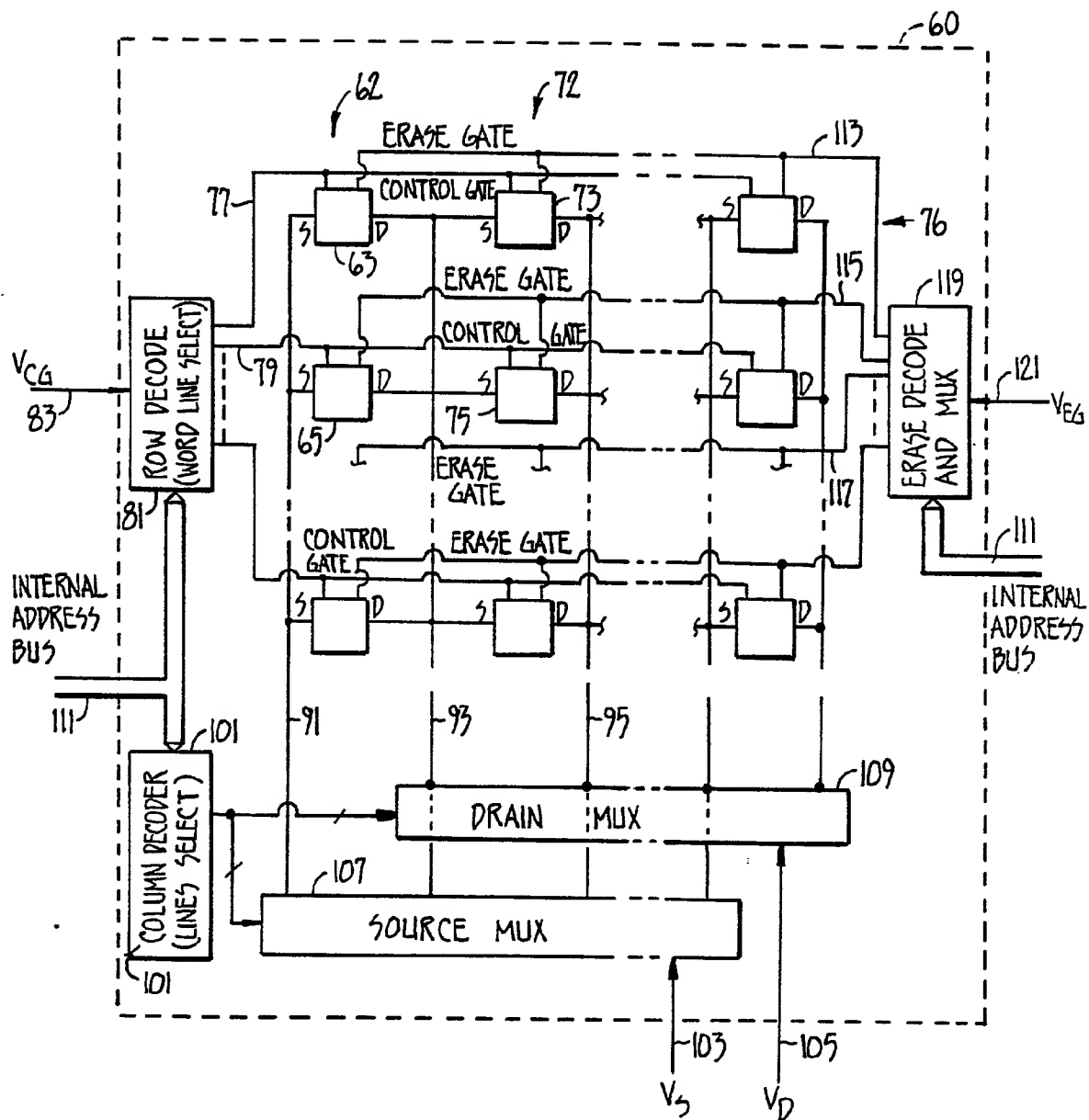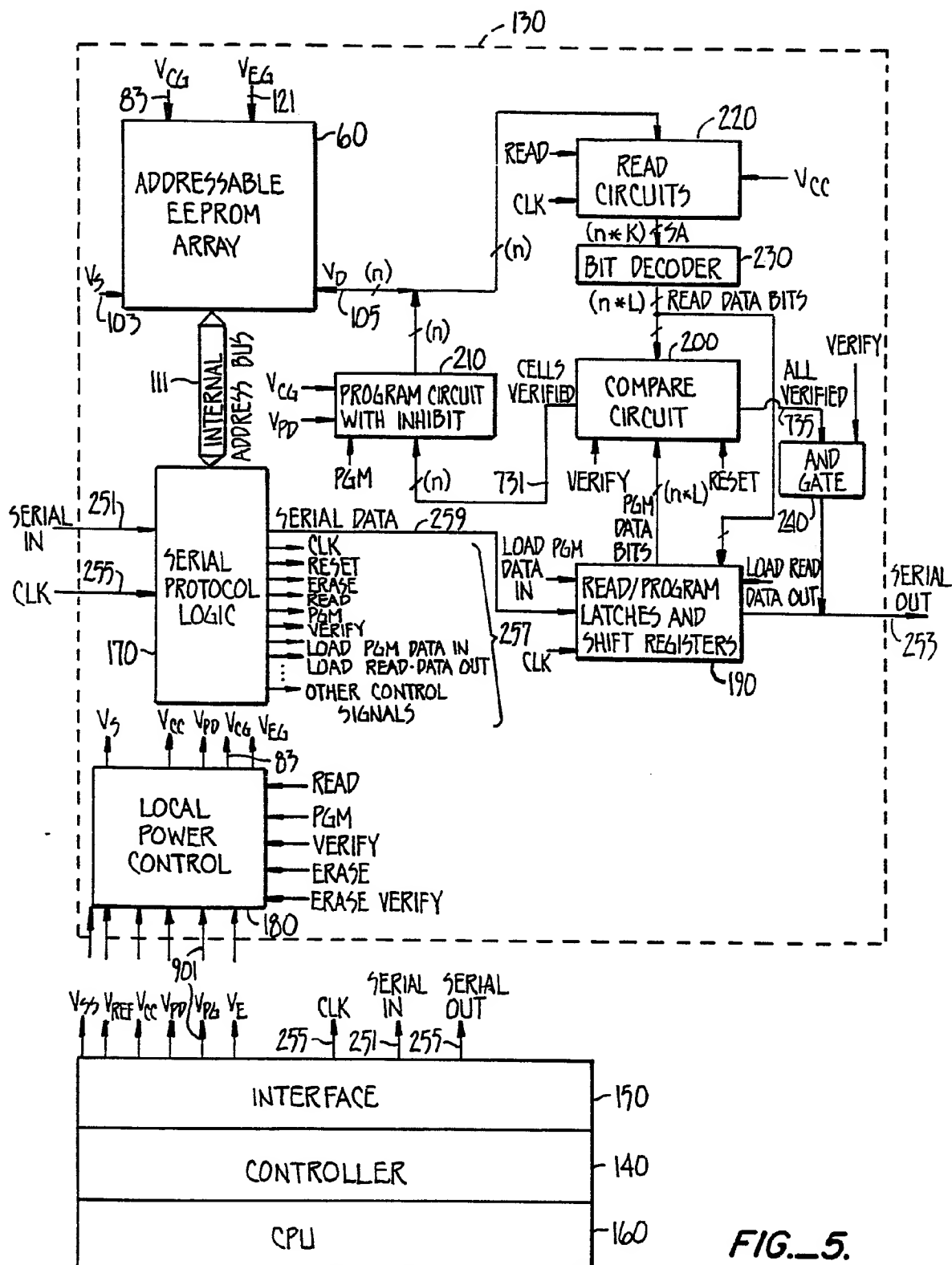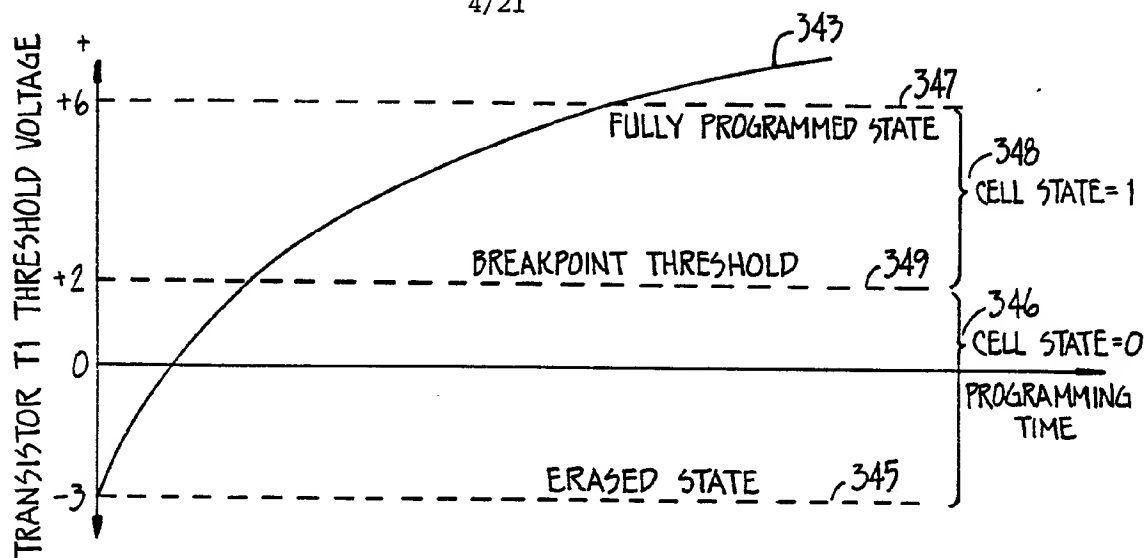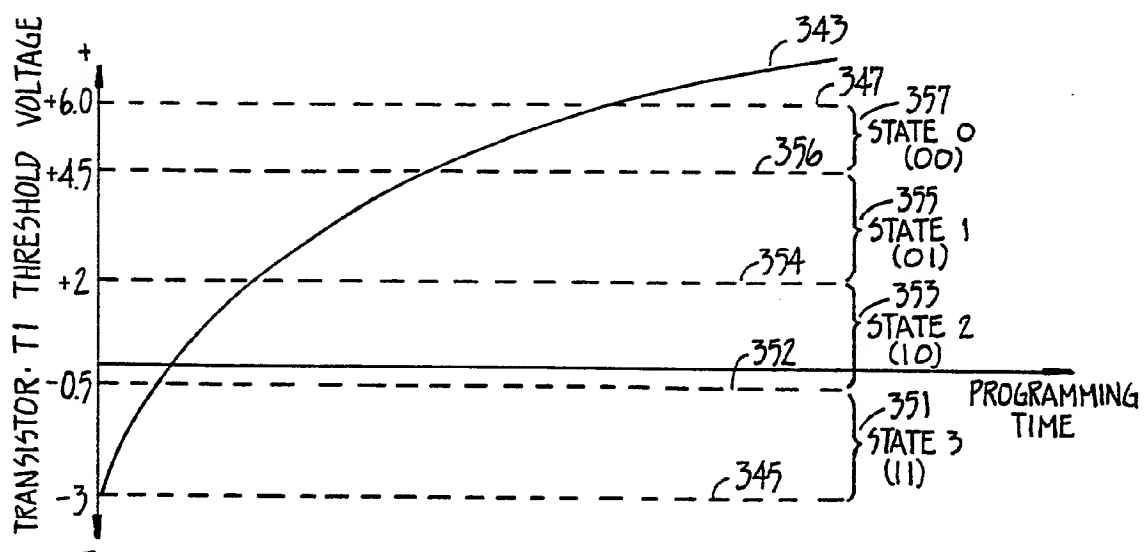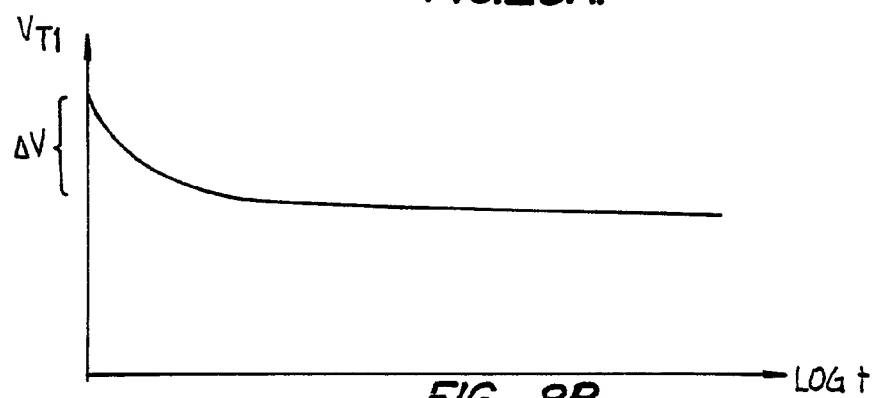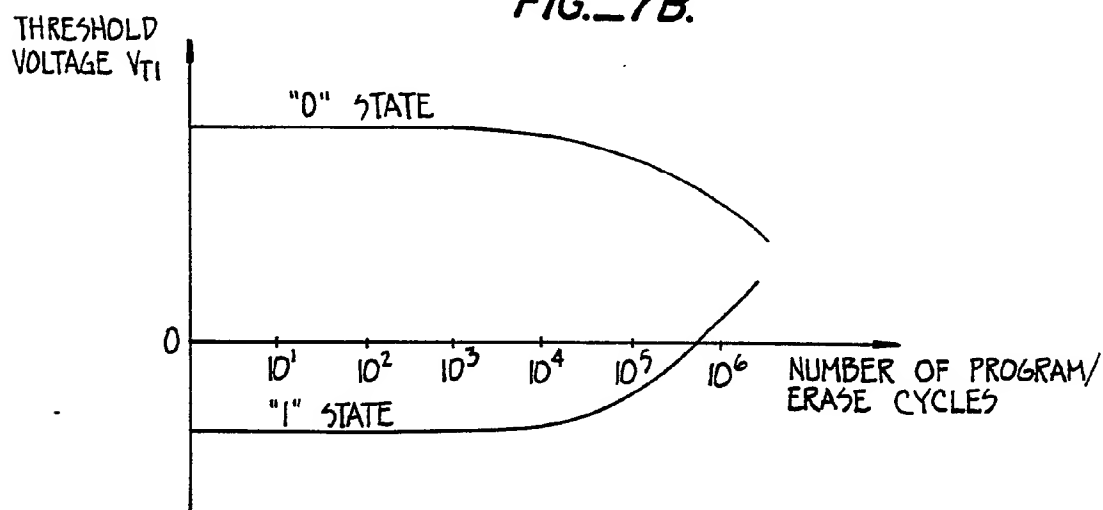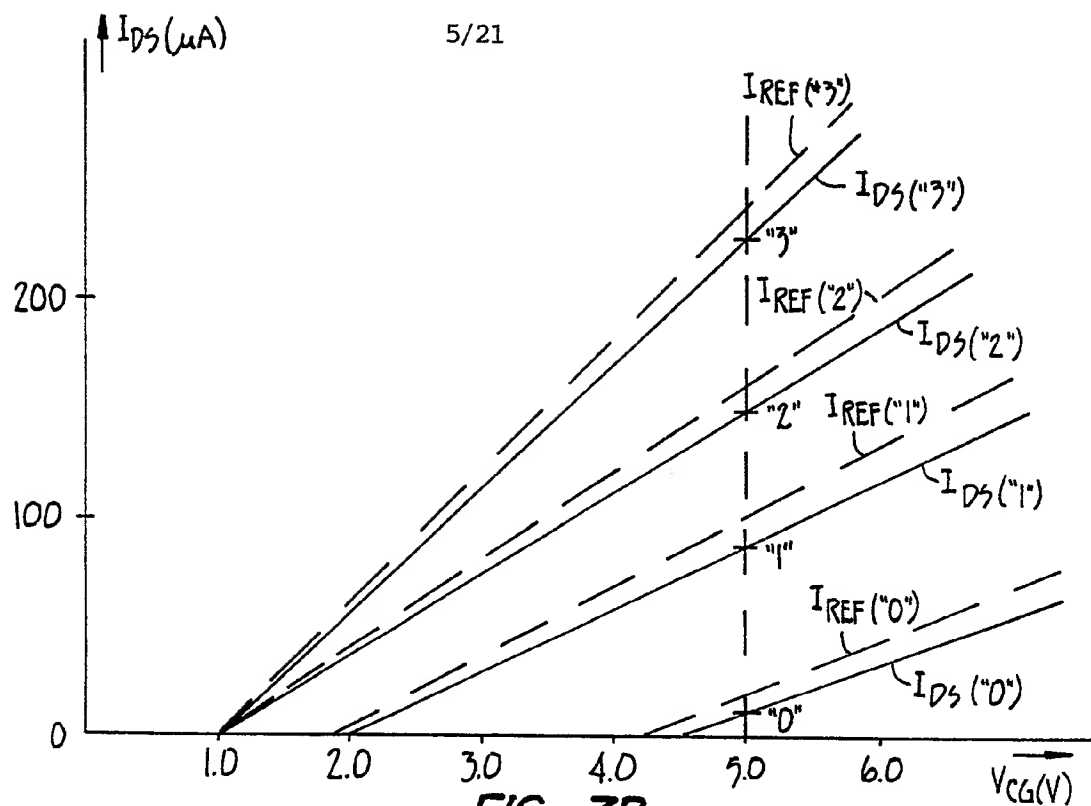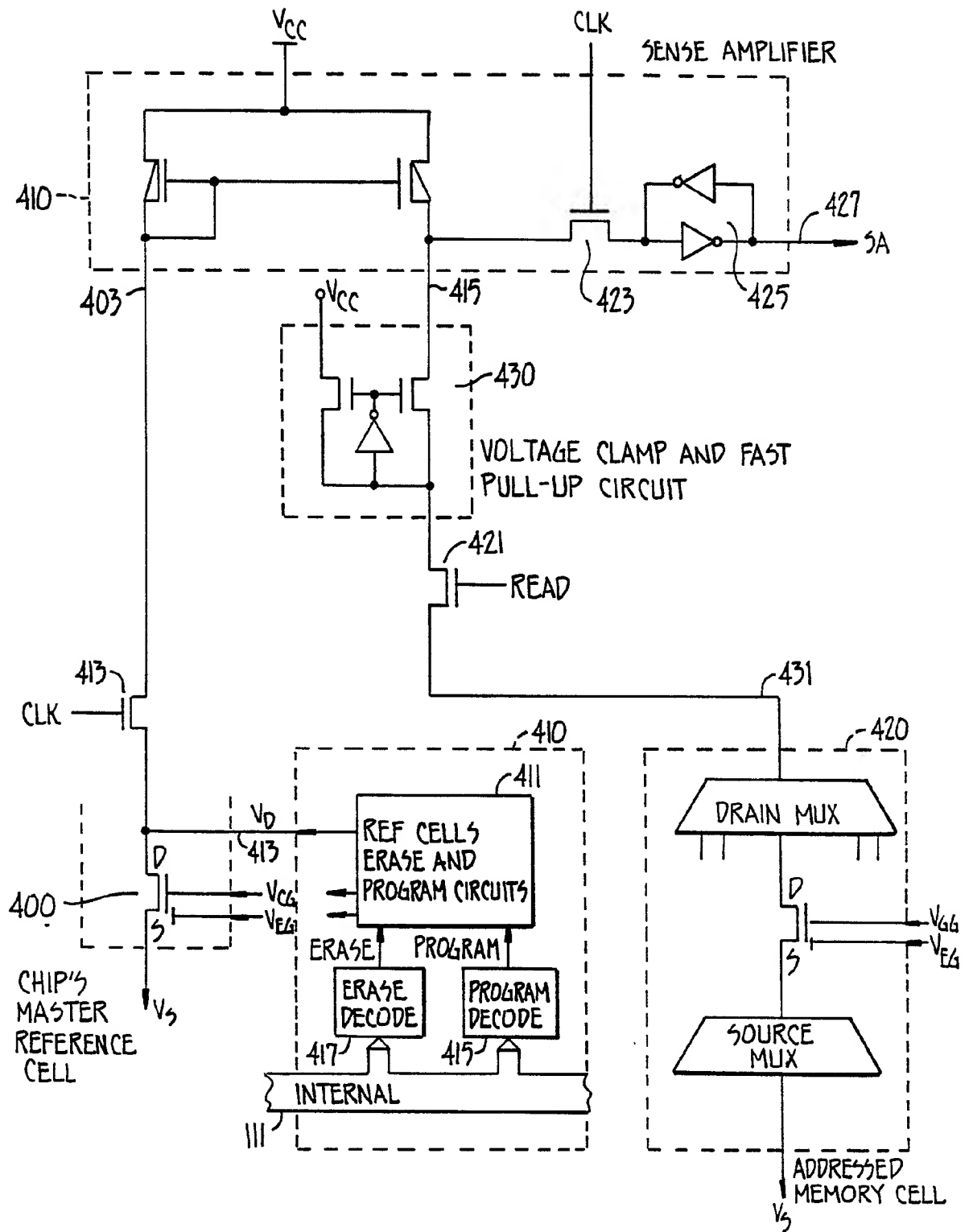
FIG._4.

FIG._5.

4/21

FIG._6.

FIG._7A.

5/21

FIG._7B.

FIG._8A.

FIG._8B.

FIG._9A.

FIG__9B.

(1) READ

(2) CLK 1

(3) CLK 2

(4) CLK K

(5) SA1

(6) SA2

(7) SAK

(8) BITS 1-L

FIG._9C.

MASTER REF. CELLS ~507

| LOCAL SECTOR READ CIRCUIT | SECTOR #1 | ~501 |

| LOCAL SECTOR READ CIRCUIT | SECTOR #2 | ~503 |

513

| LOCAL SECTOR READ CIRCUIT | SECTOR #N | ~505 |

FIG._10.

FIG._9D.

FIG._9E.

FIG._9F.

FIG.__9G.

FIG.__9H.

FIG._9I.

FIG._11.

14/21

V_CC   CLK

440 ——— SENSE
         AMPLIFIER   —→ SA

441 —   — 457

LOCAL REF   527   521   521
PROGRAM/   ⊗   READ — ⊗   ⊗   ⊗
ERASE
VERIFY

                529   525   523
CHIP'S   SECTOR'S   SECTOR'S
MASTER   —CLK   LOCAL   —CLK   ADDRESSED
REFERENCE   REFERENCE   MEMORY
CELLS   CELLS   CELLS

FIG._12A.

V_CC   CLK

                    SENSE   440
                    AMPLIFIER   —→ SA

        441 —   — 457

                    READ I — ⊗ 533   READ — ⊗ 535

        531   525   523
BIASED CHIP'S   SECTOR'S   SECTOR'S
MASTER   —CLK   LOCAL   —CLK   ADDRESSED
REFERENCE   REFERENCE   MEMORY
CELLS   CELLS   CELLS
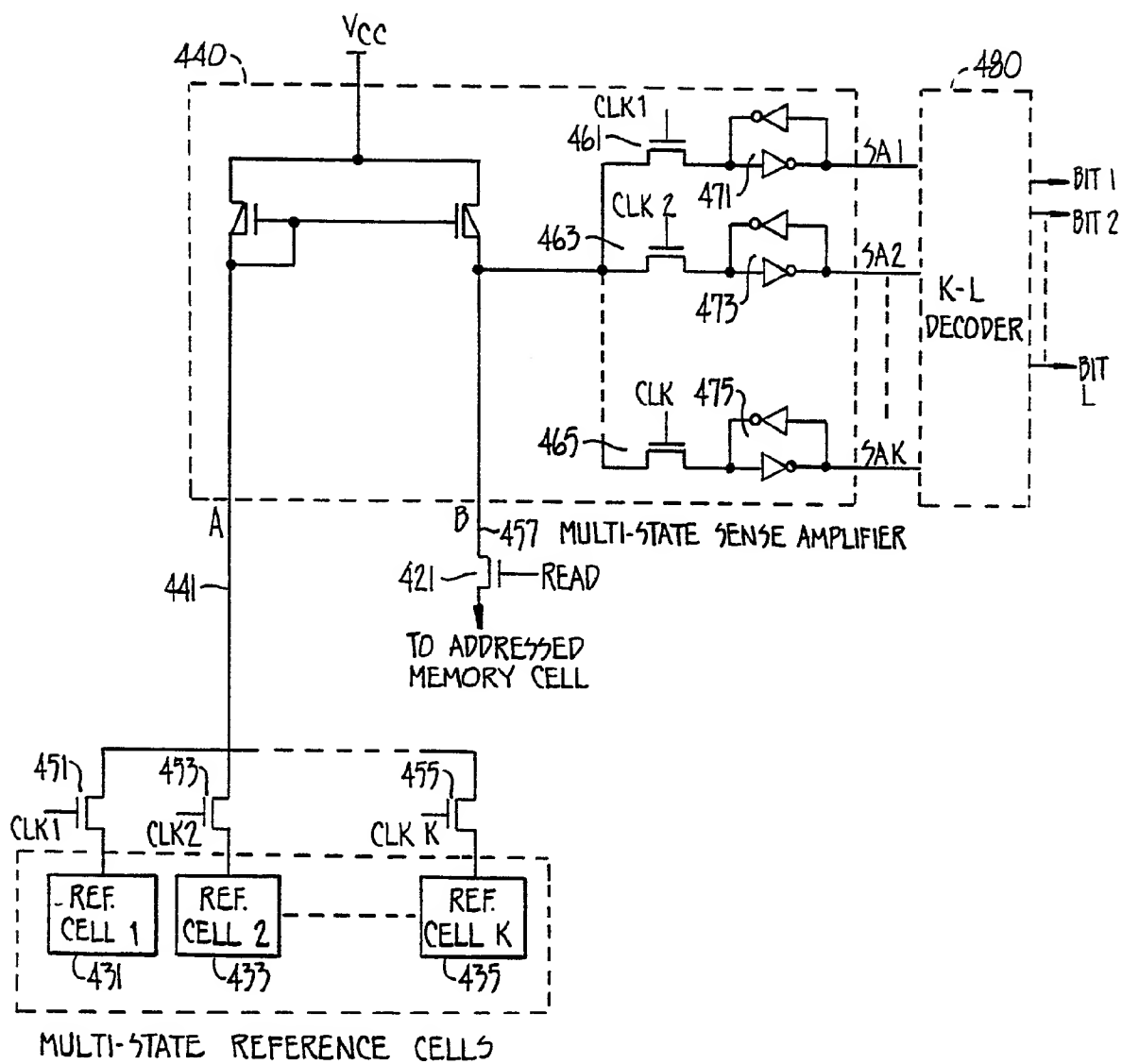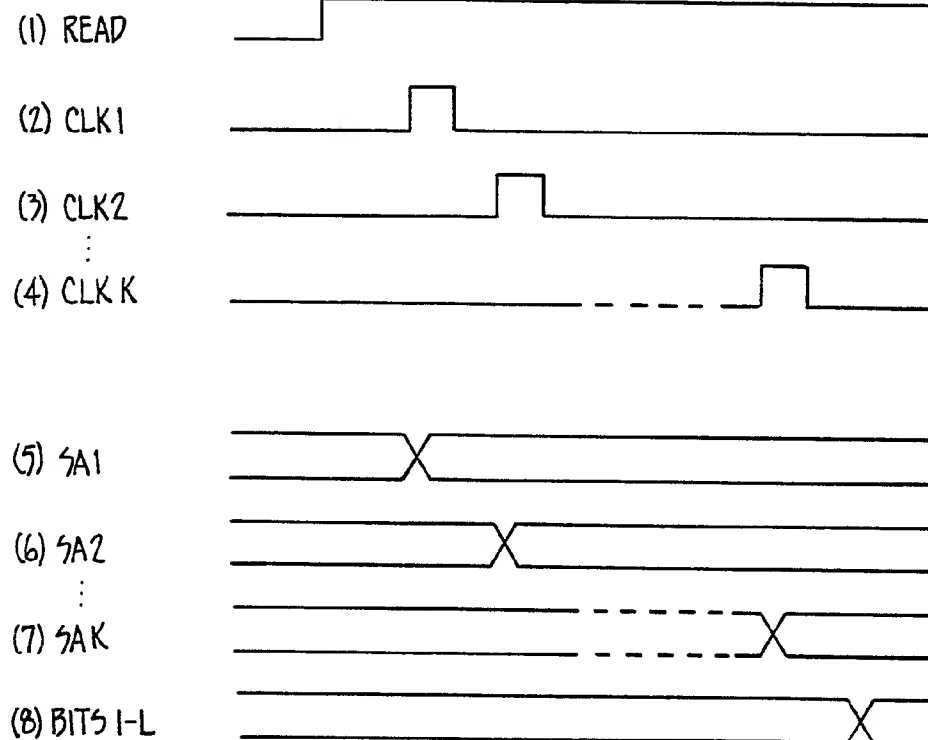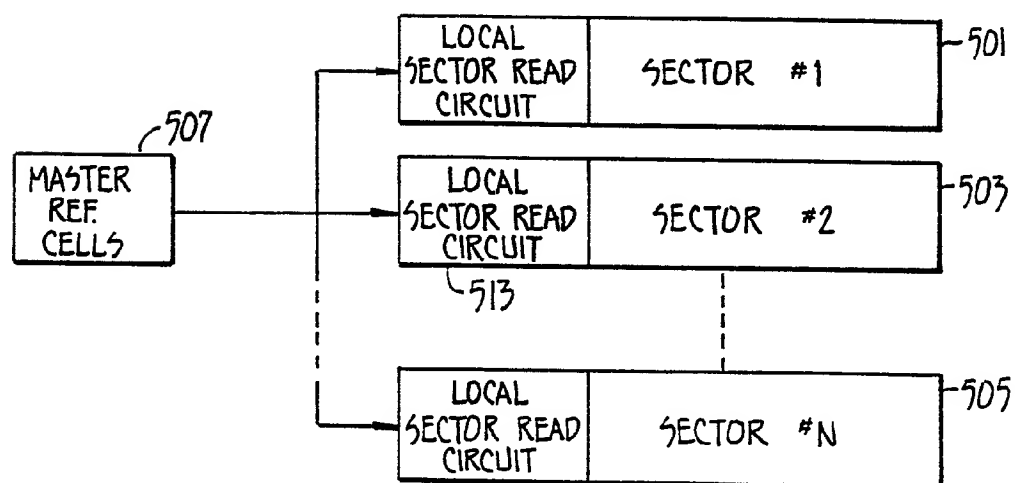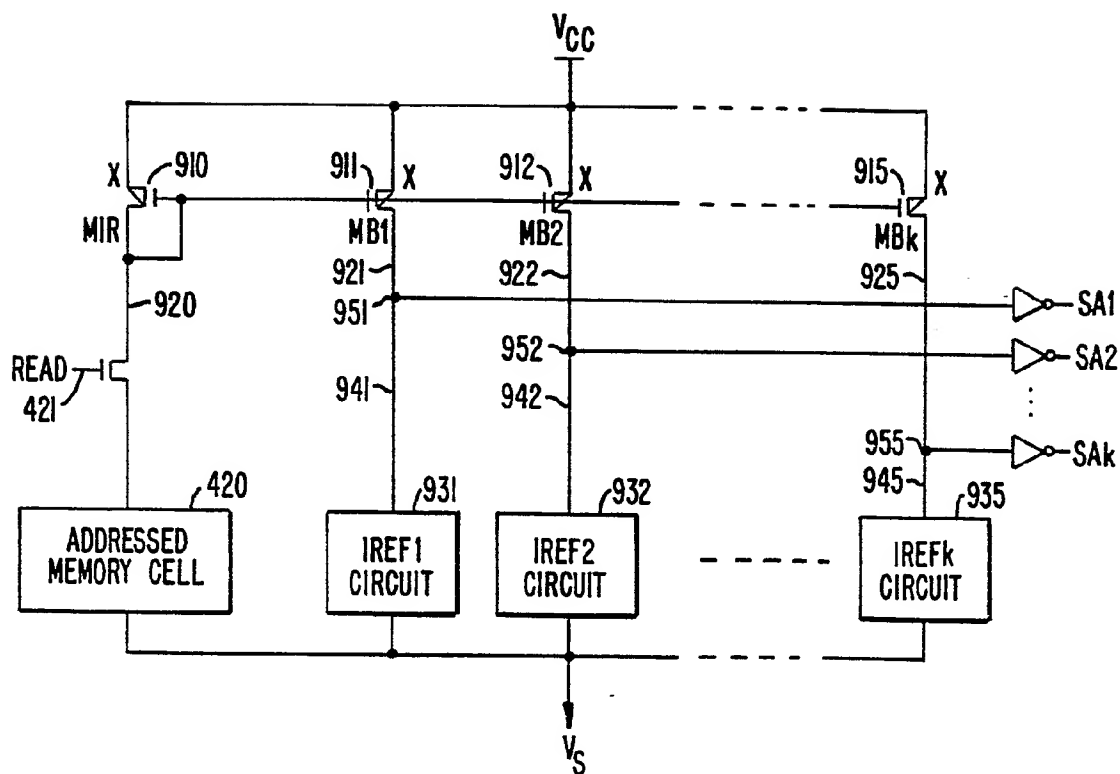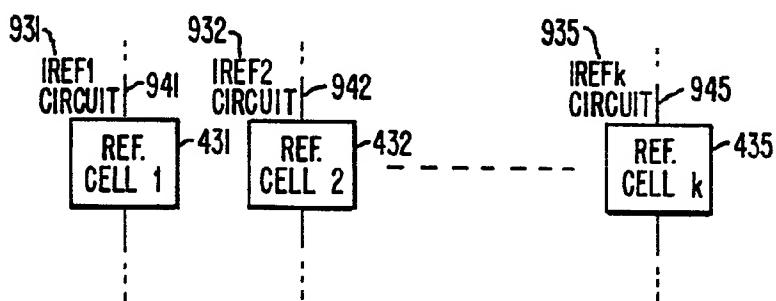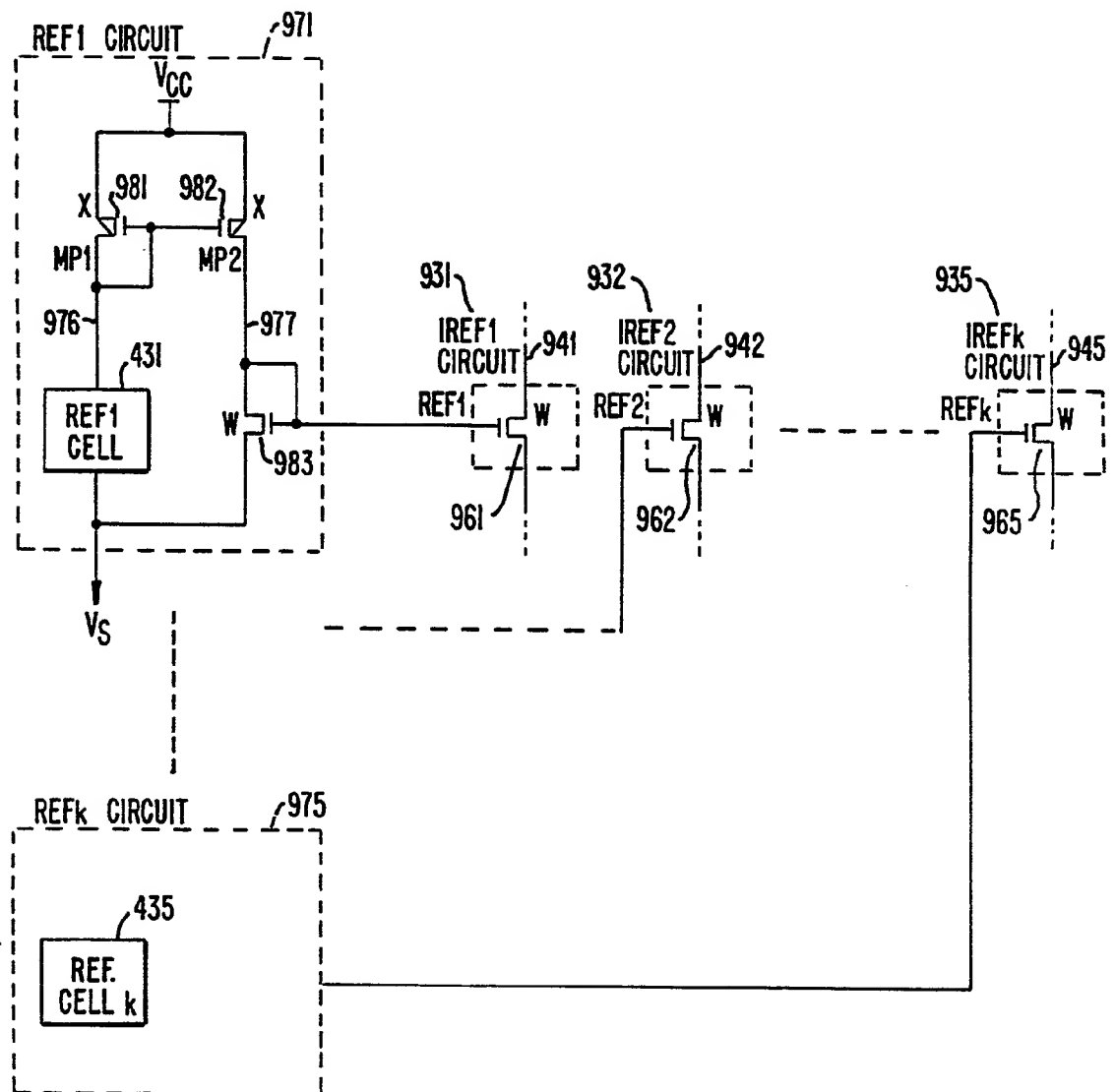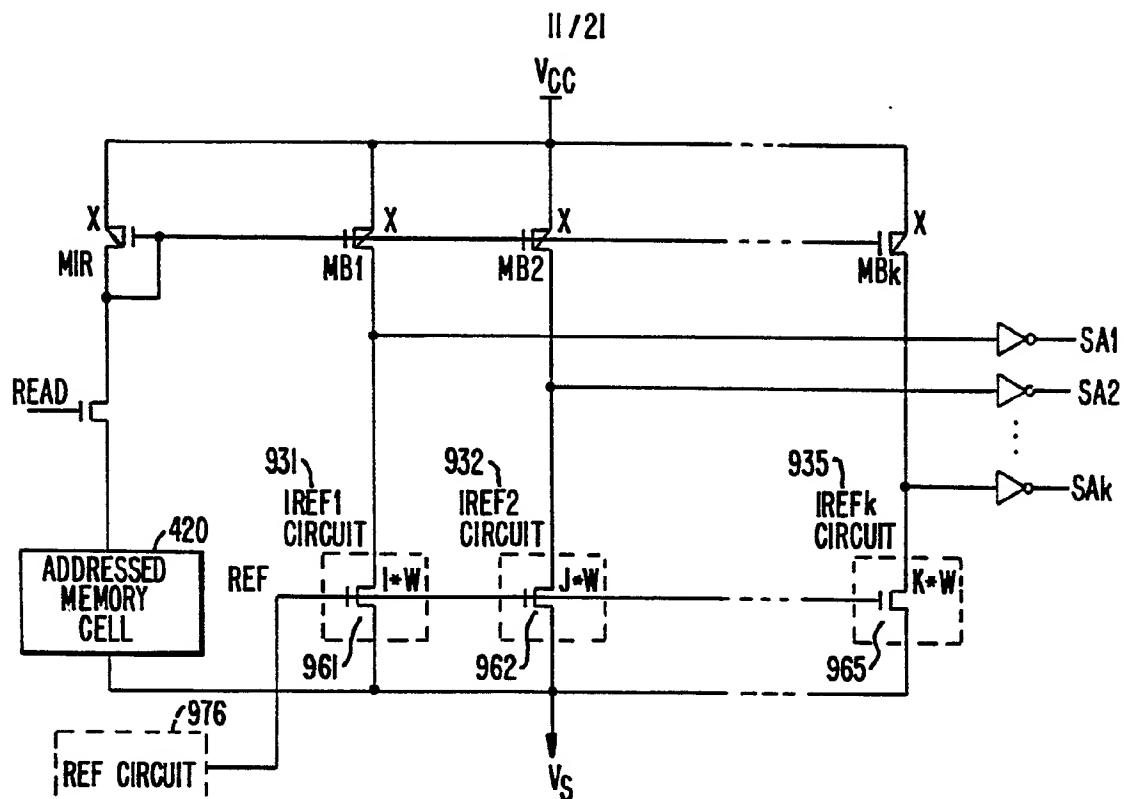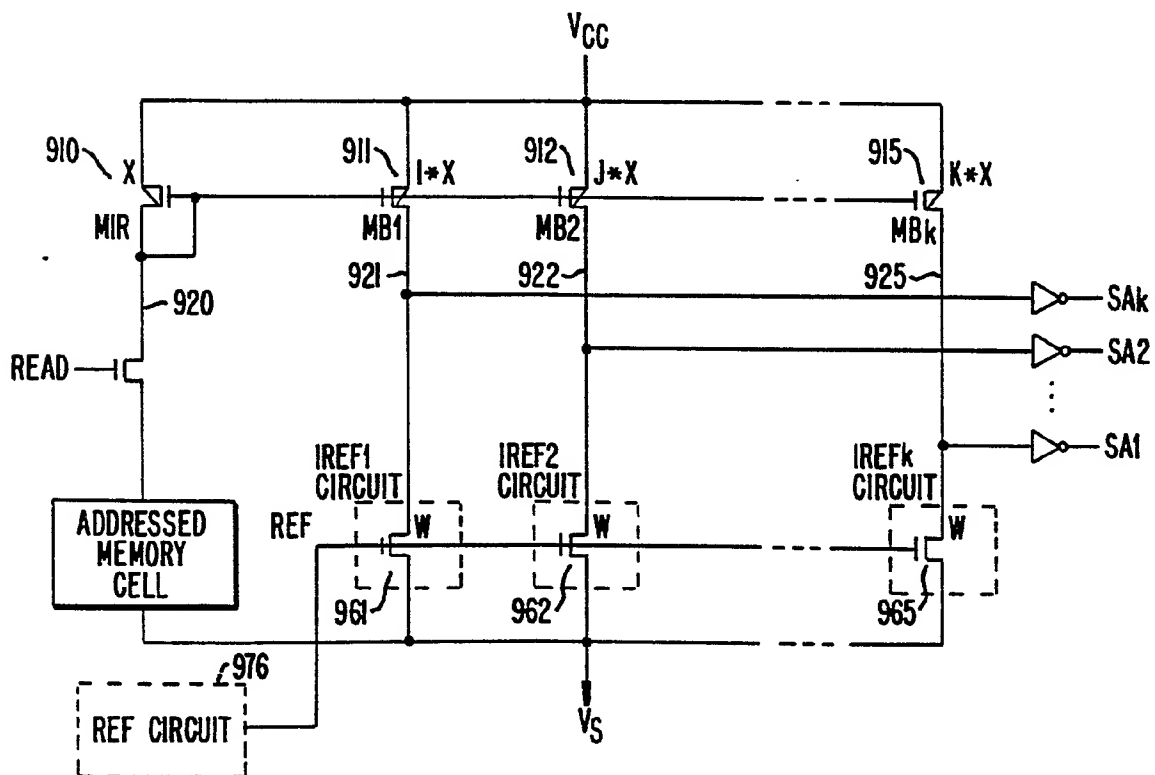
FIG._13A.
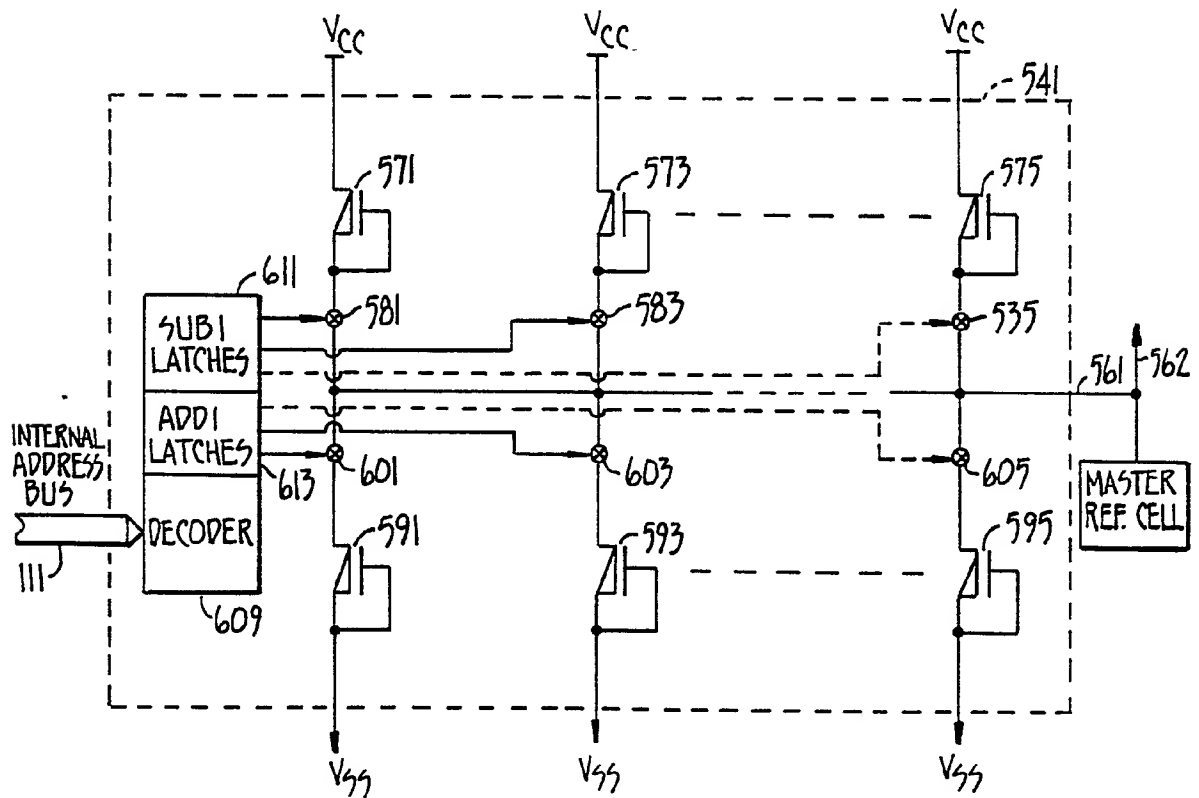
SUBSTITUTE SHEET
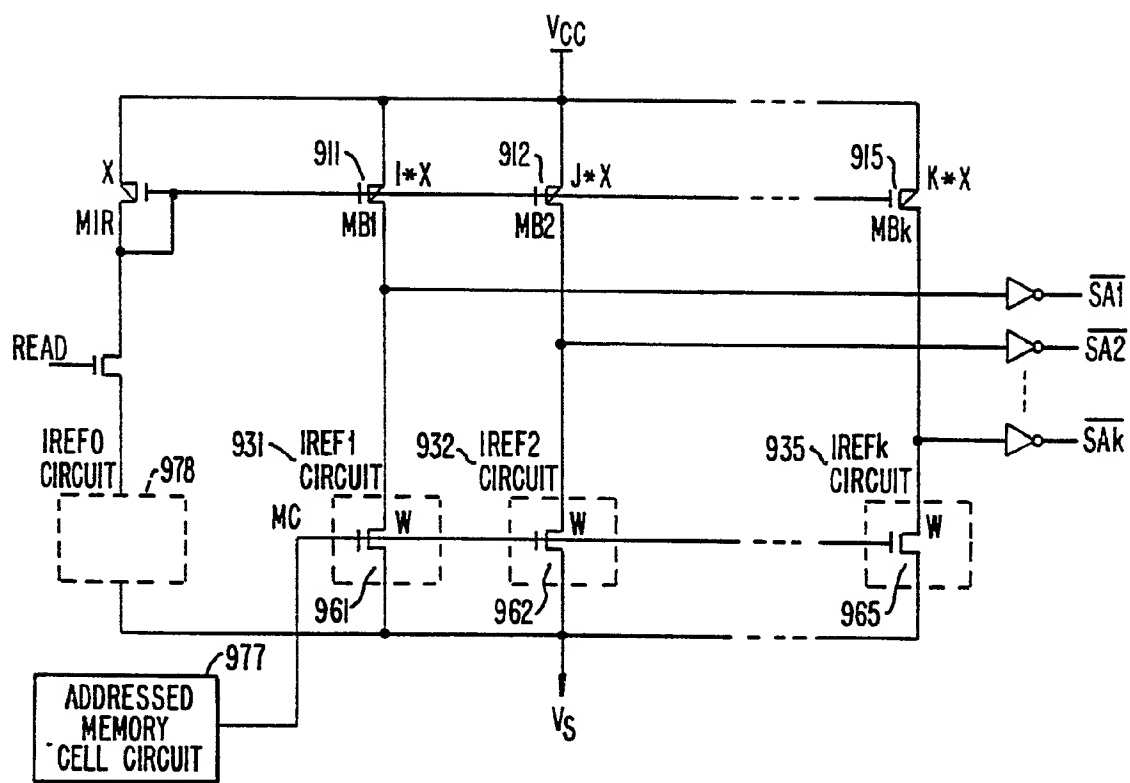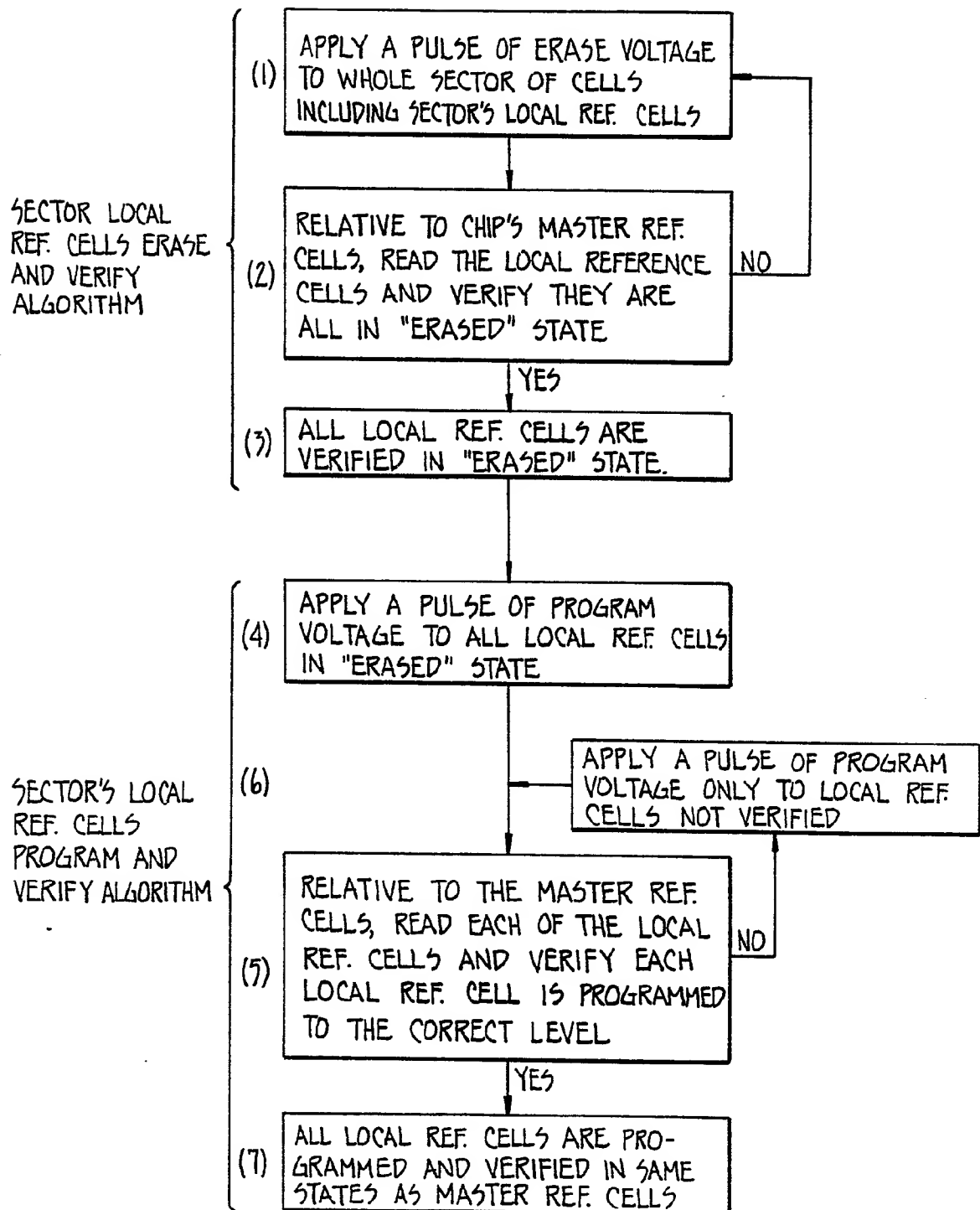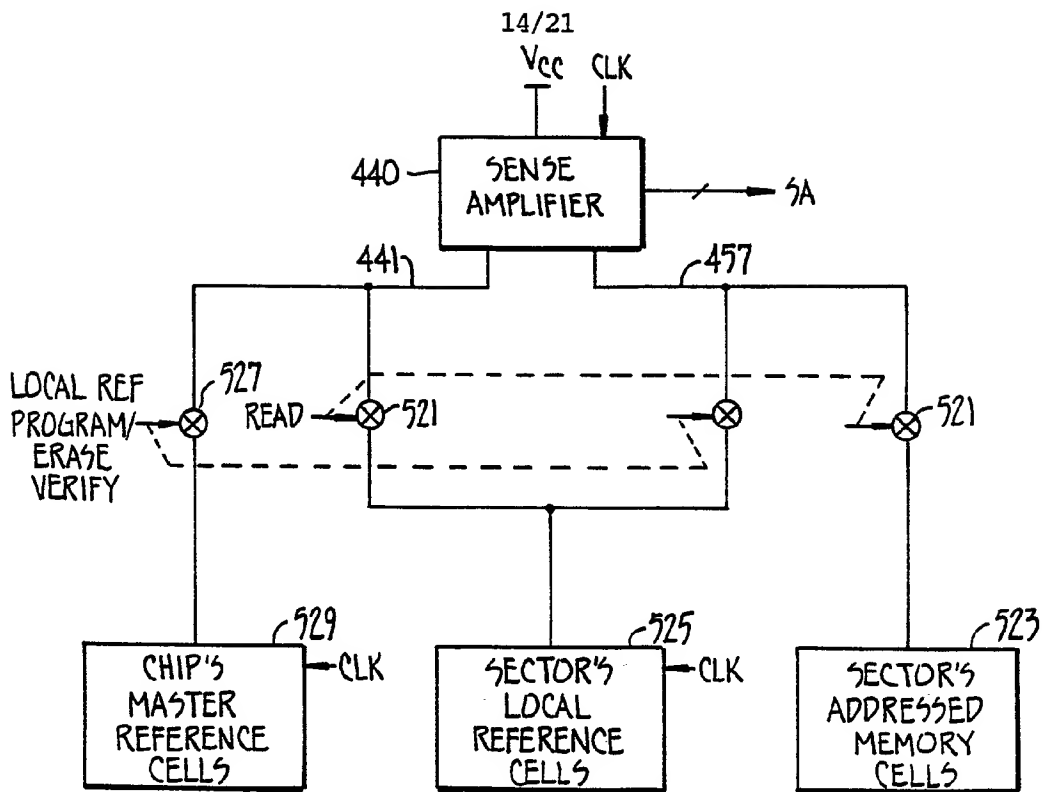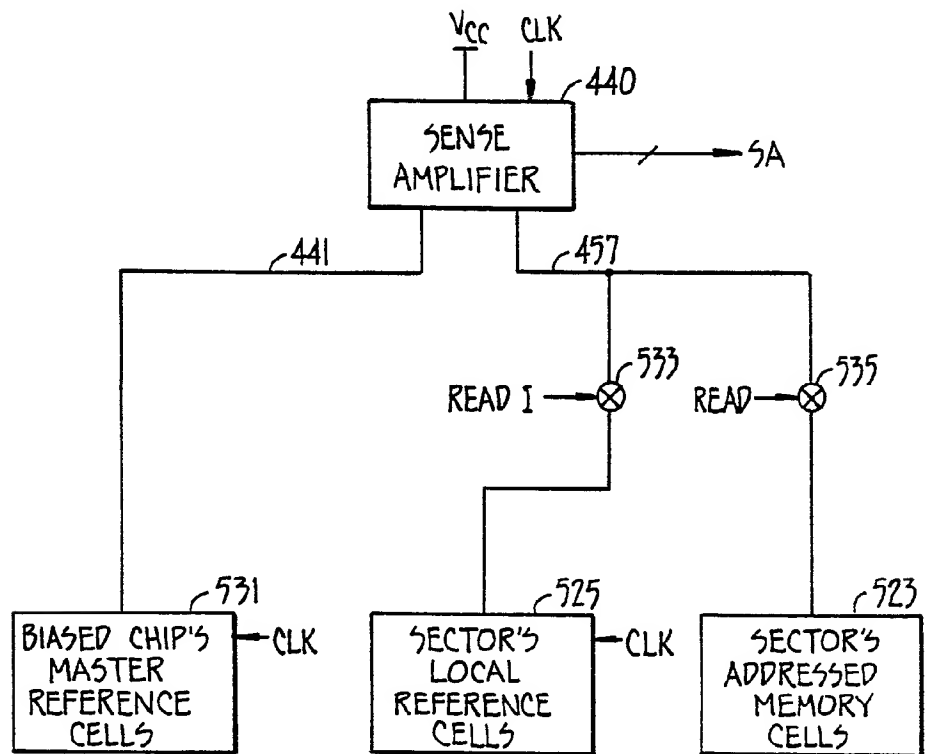
TO SENSE AMPLIFIER
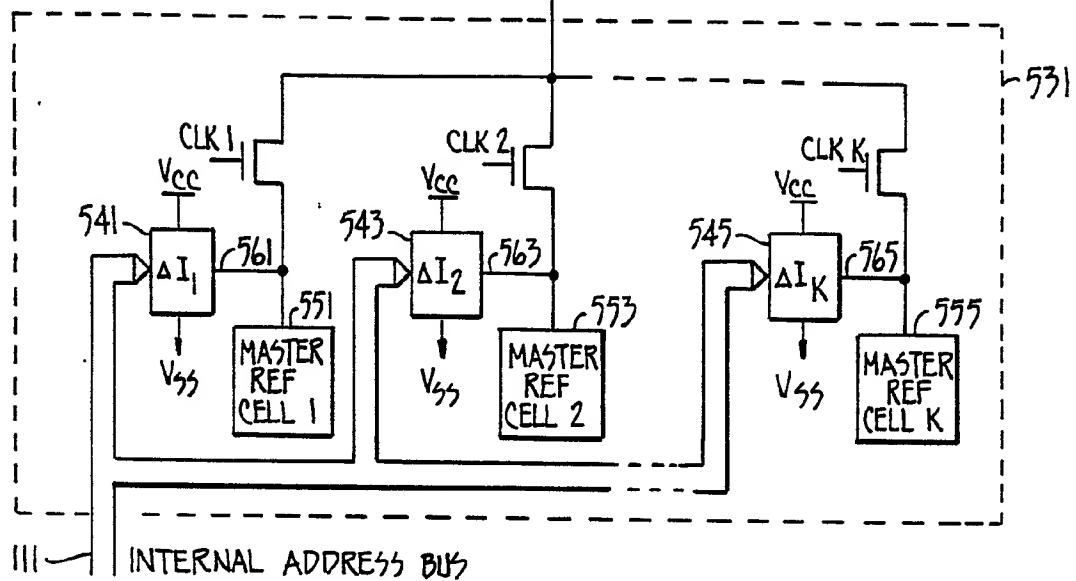
FIG._13B.

FIG._13C.

LOCAL REF. CELLS ARE PREVIOUSLY PROGRAMMED
AND VERIFIED IN SAME STATES AS MASTER REF. CELLS

RELATIVE TO THE LOCAL REF. CELLS,
READ THE ADDRESSED CELLS

FIG._12B.

(1) LOCAL REF. CELLS ARE PREVIOUSLY
     PROGRAMMED AND VERIFIED IN SAME
     STATES AS MASTER REF. CELLS

(2) RELATIVE TO THE LOCAL REFERENCE
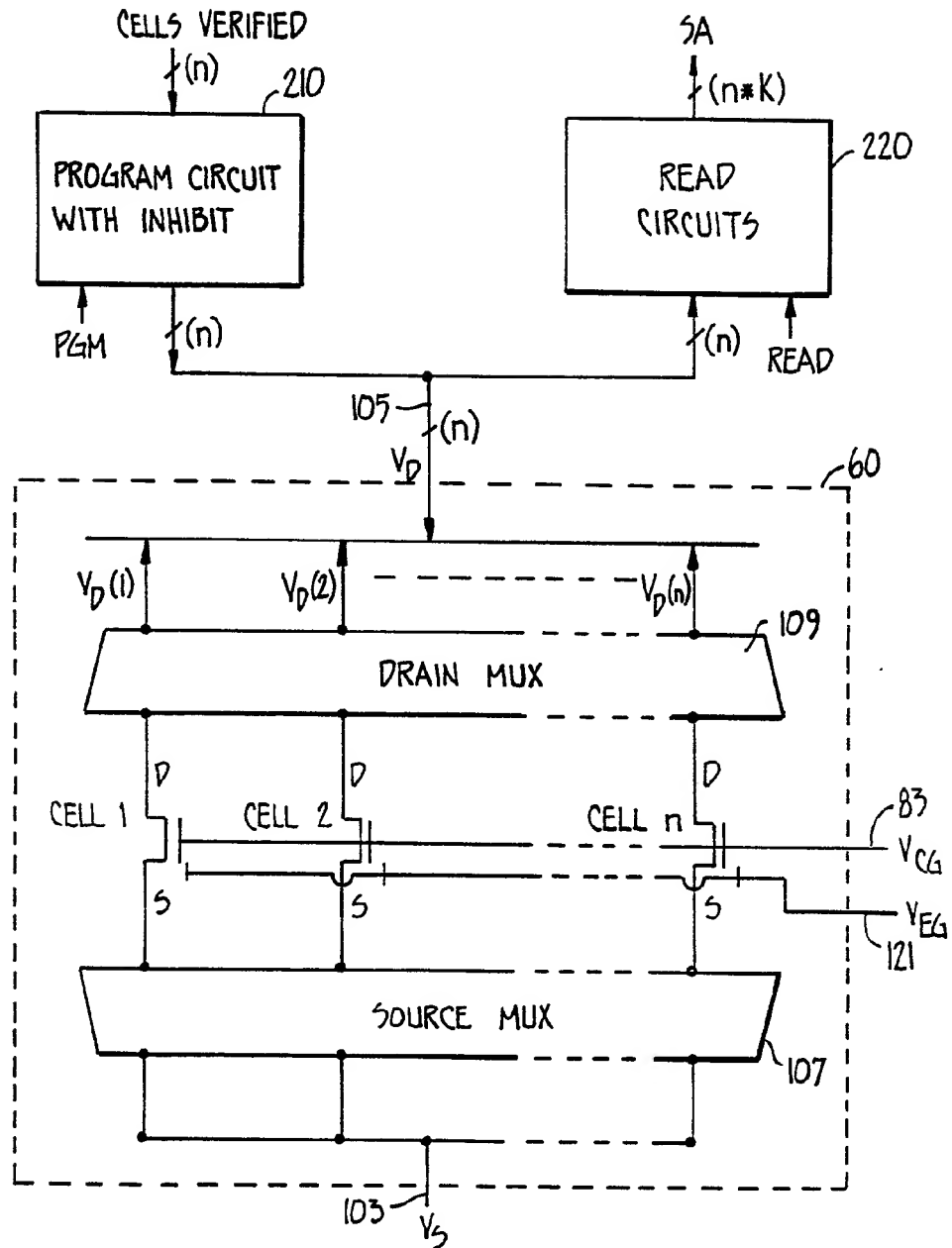   CELLS READ THE MASTER REF. CELLS

(3) DETERMINE THE DIFFERENCES, IF ANY
     AND BIAS. THE MASTER REF CELLS'
     CURRENTS SUCH THAT THE SAME
     READING IS OBTAINED RELATIVE TO
     THE BIASED MASTER REF. CELLS
     AS RELATIVE TO THE LOCAL
     REF. CELLS

(4) RELATIVE TO THE BIASED MASTER
     REF. CELLS, READ THE ADDRESSED CELLS

FIG._13D.

READ/PROGRAM DATA PATHS FOR n CELLS IN PARALLEL

FIG_14.

(1) APPLY SECTOR ERASE AND VERIFY ALGORITHM

(2) APPLY SECTOR'S REF. CELLS PROGRAM AND VERIFY ALGORITHM

(3) LATCH DATA FOR PROGRAMMING ADDRESSED CELLS

(6) APPLY A PULSE OF PROGRAM VOLTAGE ONLY TO ADDRESSED CELLS NOT VERIFIED

(4) APPLY READ ALGORITHM FOR DATA STORED IN ADDRESSED CELLS

(5) VERIFY READ DATA = PROGRAM DATA FOR ALL ADDRESSED CELLS   NO

YES

(7) ALL ADDRESSED CELLS ARE PROGRAMMED AND VERIFIED

PROGRAM ALGORITHM

FIG._15.
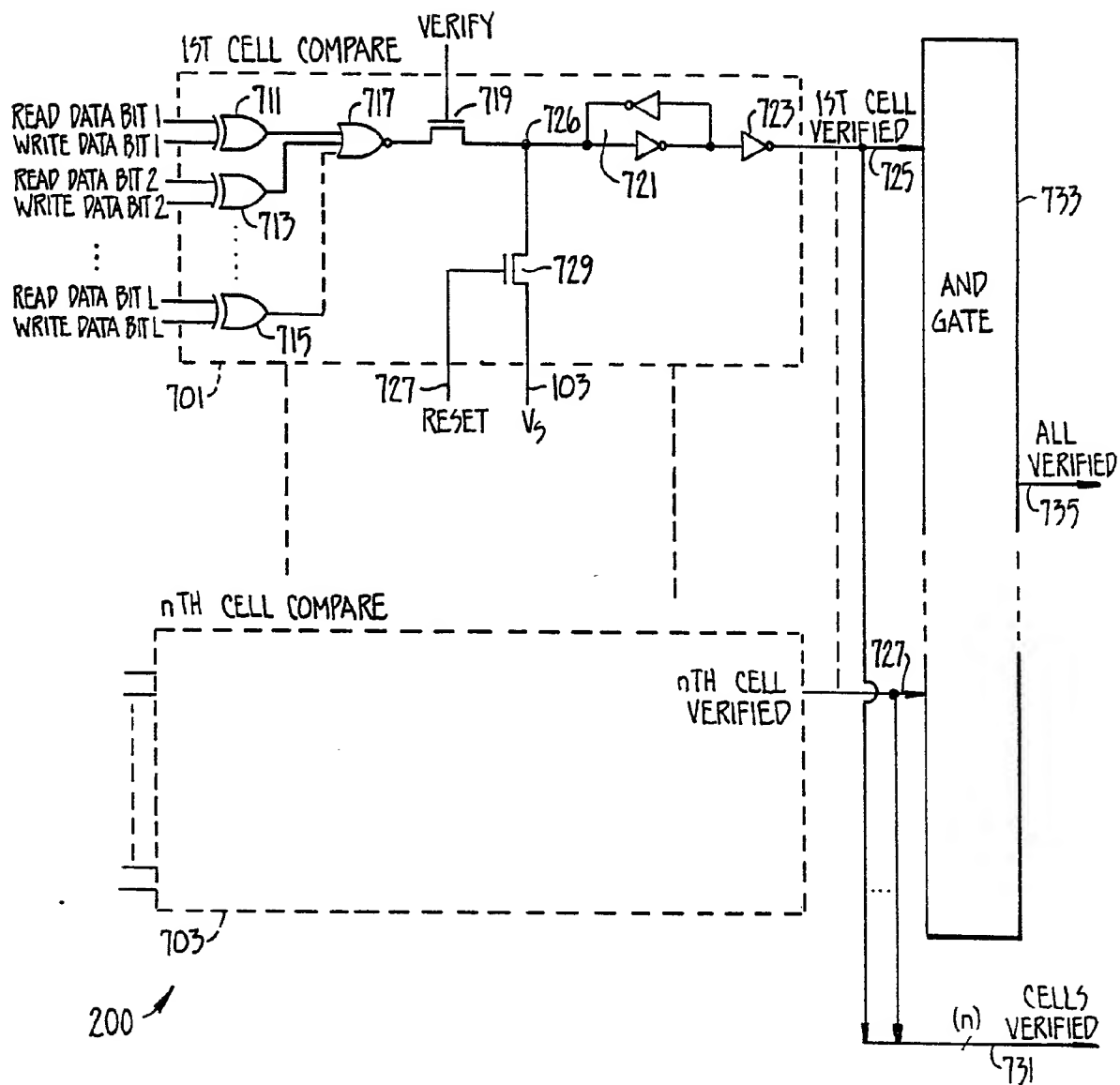
FIG._16.
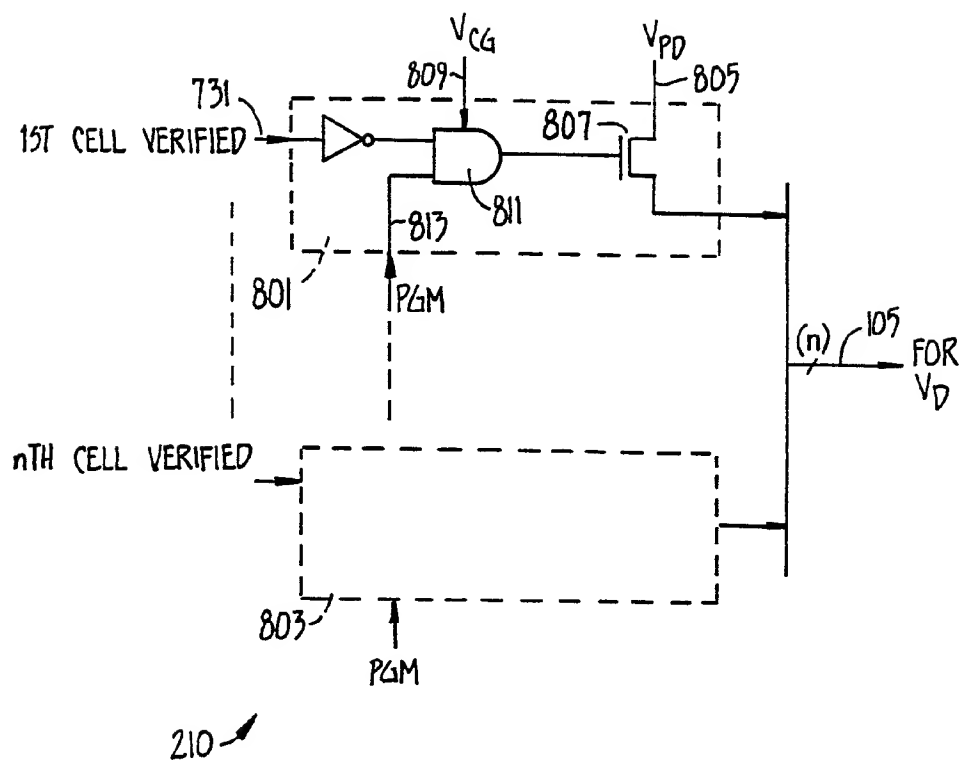
FIG_17.

	SELECTED CONTROL GATE V_{CG}	DRAIN V_D	SOURCE V_S	ERASE GATE V_{EG}
READ	V_{PG}	V_{REF}	V_{SS}	V_E
PROGRAM	V_{PG}	V_{PD}	V_{SS}	V_E
PROGRAM VERIFY	V_{PG}	V_{REF}	V_{SS}	V_E
ERASE	V_{PG}	V_{REF}	V_{SS}	V_E
ERASE VERIFY	V_{PG}	V_{REF}	V_{SS}	V_E

TABLE 1

(TYPICAL VALUES)	READ	PROGRAM	PROGRAM VERIFY	ERASE	ERASE VERIFY
V_{PG}	V_{CC}	12V	$V_{CC}+\delta V$	V_{CC}	$V_{CC}-\delta V$
V_{CC}	5V	5V	5V	5V	5V
V_{PD}	V_{SS}	8V	8V	V_{SS}	V_{SS}
V_E	V_{SS}	V_{SS}	V_{SS}	20V	V_{SS}
UNSELECTED CONTROL GATE	V_{SS}	V_{SS}	V_{SS}	V_{SS}	V_{SS}
UNSELECTED BIT LINE	V_{REF}	V_{REF}	V_{REF}	V_{REF}	V_{REF}

$V_{SS}=0V,$ $V_{REF}=1.5V,$ $\delta V=0.5V-1V$

TABLE 2

SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) 3

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): G11C 7/00, 11/00, 16/00; G01R 31/28
U.S. CL.: 365/185; 104; 201; 189.07

II. FIELDS SEARCHED

Minimum Documentation Searched 4

Classification System	Classification Symbols
U.S.	365/96,104,185,168,189.07,189.09,201,184 371/21.4

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched 5

III. DOCUMENTS CONSIDERED TO BE RELEVANT 14

Category *	Citation of Document, 16 with indication, where appropriate, of the relevant passages 17	Relevant to Claim No. 18
X / Y	US, A, 4,799,195 (IWAHASHI ET AL.) 17 January 1989, See columns 4-7.	1,13 / 8,16,27,46
X / Y	US, A, 4,733,394 (GIEBEL) 22 March 1977, See column 2, lines 26-42, column 4.	1,13,36,46,51, 52,53 / 54,8,16,27,46
A	US, A, 4,612,629 (HARARI) 16 September 1986, See column 5 line 60 to column 10 line 20. See also column 14 line 42 to column 15 line 41.	ALL
A	US, A, 4,252,059 (BELL ET AL.) 24 February 1981, See column 2, lines 20-43.	59-63,65-87
A	US, A, 4,460,982 (GEE ET AL.) 17 JULY 1984, See column 6, lines 5-24.	59-63,65-87
A	US, A, 4,809,231 (SHANNON ET AL.) 28 February 1989, See column 2, lines 57-68 to column 3, lines 1-6.	59-63,65-87

* Special categories of cited documents: 15

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search 2	Date of Mailing of this International Search Report 2
31 JULY 1990	14 AUG 1990

International Searching Authority 1	Signature of Authorized Officer 20
ISA/US	ALYSSA BOWLER

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers   , because they relate to subject matter [1] not required to be searched by this Authority, namely:

2. ☐ Claim numbers .... , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [1], specifically:

3. ☐ Claim numbers_____ , because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

I.   Claims 1-58, 64 drawn to a memory cell array.

II.   Claims 59-63 and 65-87 drawn to a current mirror test circuit.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.   Telephone Practice

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (supplemental sheet (2) (Rev. 4-90)